

Optimal and Scalable Methods to Approximate the Solutions of Large-Scale Bayesian Problems: Theory and Application to Atmospheric Inversions and Data Assimilation

N. Bousserez and Daven K. Henze

2016/09/19

Abstract

This paper provides a detailed theoretical analysis of methods to approximate the solutions of high-dimensional ($> 10^6$) linear Bayesian problems. An optimal low-rank projection that maximizes the information content of the Bayesian inversion is proposed and efficiently constructed using a scalable randomized SVD algorithm. Useful optimality results are established for the associated posterior error covariance matrix and posterior mean approximations, which are further investigated in a numerical experiment consisting of a large-scale atmospheric tracer transport source-inversion problem. This method proves to be a robust and efficient approach to dimension reduction, as well as a natural framework to analyze the information content of the inversion. Possible extensions of this approach to the non-linear framework in the context of operational numerical weather forecast data assimilation systems based on the incremental 4D-Var technique are also discussed, and a detailed implementation of a new Randomized Incremental Optimal Technique (RIOT) for 4D-Var algorithms leveraging our theoretical results is proposed.

1 Introduction

The Bayesian approach to inverse problems consists of updating a prior probability distribution of a quantity of interest conditioned on some physically-related observations. The conditioned probability distribution is called the posterior distribution. The Bayesian framework has been widely adopted to solve geophysical problems. For large-scale non-linear problems, such as those encountered in atmospheric modeling, sampling techniques (e.g., Markov Chain Monte-Carlo) to estimate the posterior distribution require prohibitively numerous integrations of the forward model that relates the inferred variables to the observations (Tarantola, 2005). Alternatively, when the forward model is linear and the probability distributions for the prior and the observations are Gaussian, the posterior probability distribution is Gaussian and can be fully characterized by its mean (i.e., the maximum-likelihood) and its error covariance matrix, for which analytical expressions are available. However, explicitly calculating the posterior error covariance matrix remains a challenging task in inverse problems for which the dimensions of the matrix are very large (Bousserez *et al.*, 2015). As an example, the typical number of optimized variables in data assimilation (DA) systems for numerical weather prediction (NWP) is $\sim 10^8$, which corresponds to a posterior error covariance matrix of dimension $10^8 \times 10^8$. In such cases, the posterior error covariance matrix cannot be explicitly represented in computer memory, and appropriate low-rank approximations are needed to extract useful information. Low-rank estimations of the posterior error covariance matrix are also useful to compute other quantities of interest such as the model resolution matrix

and the Degree Of Freedom for Signal (DOFS), which characterize the information content of the inversion (Rodgers, 2000; Tarantola, 2005).

The variational approach to solving large-scale inverse problems employs tangent-linear and adjoint models with iterative gradient-based optimization algorithms to compute the maximum-likelihood of the posterior distribution. The potential of state-of-the-art optimization algorithms to provide posterior error covariance estimates as a by-product of the minimization have long been recognized (e.g., Thacker (1989); Rabier and Courtier (1992); Nocedal and Wright (2006); Müller and Stavrakou (2005)). For large-scale problems, such optimization algorithms are usually halted before full convergence, effectively only approximating the solution. Although the convergence properties of these approximations have been investigated in previous numerical experiments (e.g., Bousserez *et al.* (2015)), a theoretical analysis of their optimality with respect to the information content of the inversion has yet to be performed. Another approach to make large-scale inverse problems tractable is the use of ensembles to approximate the error covariances of the system. Such methods, which can be either stochastic (e.g., the Ensemble Kalman Filter (EnKF)) or deterministic (i.e., square-root formulations such as the Ensemble Adjusted Kalman Filter (EAKF)) have the advantage that they do not require the use of an adjoint model. However, the small number of ensembles used results in severe rank-deficiencies for the associated error covariance matrices. Sophisticated filtering localization techniques that help mitigate this sampling noise are the subject of intense research activities in ensemble-based DA (e.g., Ménétrier *et al.* (2015); Anderson and Lei (2013)).

Besides implicit (i.e., incomplete variational minimizations) and explicit (i.e., ensemble-based) low-rank approximations, another approach to approximate the solution of large-scale Bayesian problems consists of performing a prior dimensional reduction of the system. In the context of linear problems, such methods can allow one to explicitly form the matrices associated with the inverse problem and to analytically compute the posterior mean and posterior error covariance. In the atmospheric inversion community, several studies have focused on designing objective methods to construct reduced spaces, so as to optimize some criteria related to the information content of the problem. In Bocquet *et al.* (2011), a rigorous multi-scale approach to dimension reduction is proposed, wherein an optimal aggregation scheme is defined by constructing a grid of tiles for which the associated reduced Bayesian problem has maximum DOFS. However, the optimization of the grid can be computationally expensive, and the method requires explicitly calculation of the Jacobian of the system, which is not feasible for high-dimensional systems. Moreover, the optimality of the solution is guaranteed only for a specific class of reductions, namely grid aggregation methods. More recently, Turner and Jacob (2015) proposed a method to construct an optimal reduced basis set of Gaussian-mixture functions to analytically solve large-scale atmospheric source inversion problems. Their algorithm consists of an incremental construction of the basis where the posterior error variances in observation space are recomputed at each iteration (i.e., for each new dimension) until a minimum total error is reached. The cost associated with computing the posterior error variances at each iteration makes this approach poorly scalable and not suitable for problems where the optimal basis needs to be constructed in a timely manner. Similarly to Bocquet *et al.* (2011), their approach also lacks generality by restricting the analysis to a specific class of basis (namely, the Gaussian-mixture functions).

Recently, Spantini *et al.* (2015) presented a detailed theoretical analysis of optimal low-rank approximations of the posterior mean and posterior error covariance matrix for linear Bayesian problems. They show that the proposed approximations are defined in the subspace that maximizes the observational constraints, which is measured as the relative gain in information in the posterior with respect to the prior information. Interestingly, this method can reconcile theoretical optimality and computational scalability, since in practice the low-rank optimal approximations can be efficiently constructed by applying matrix-free singular value decomposition (SVD) routines to the so-called prior-preconditioned Hessian of the quadratic cost function. Furthermore, when

high-performance computing is required, the use of recently developed randomized SVD methods allows to fully parallelize the algorithm and to implement a scalable approach to low-rank approximation for large-scale Bayesian problems (e.g., Bui-Thanh *et al.* (2012)).

In this paper, we provide a detailed theoretical analysis of the Bayesian approximation problem in the context of optimal projections. This approach has the advantage of producing approximations to the posterior mean and posterior error covariance matrix that are consistent with each other, i.e., they are both approximations to the full-dimensional posterior solutions and exact solutions to a projected low-rank Bayesian problem. Our mathematical developments generalize the theoretical framework of Bocquet *et al.* (2011) and allow us to construct a projection that maximizes the DOF among all low-rank projections. This maximum-DOFS projection yields posterior mean and posterior error covariance approximations similar to those proposed in Spantini *et al.* (2015), for which we provide additional interpretations and optimality results. Moreover, although Spantini *et al.* (2015) identified their optimal approximations as the solutions of a projected Bayesian problem with maximum observational information with respect to the prior, we note that they did not rigorously demonstrate this result by omitting to analyze the so-called representativeness error. This error, which quantifies the impact of the unobserved subspace in the inversion as a result of the dimension reduction, has been taken into account in our proofs. For the first time, this optimal approximation method is applied to a large-scale atmospheric-transport source inversion problem using a highly-scalable randomized SVD algorithm. Finally, we investigate new links between the maximum-DOFS approximations and preconditioned conjugate-gradient (CG) algorithms embedded in non-linear Gauss-Newton minimization methods such as incremental 4D-Var in operational DA systems for NWP. This enables us to propose an improved incremental 4D-Var algorithm leveraging both our theoretical optimality results and the efficiency of randomized SVD algorithms.

Section 2 of this paper presents the theory and formalism of the optimal low-rank projection problem and provides useful optimality results for the associated approximations of the posterior mean and posterior covariance matrices. Section 3 discusses the practical construction of the optimal approximations and describes in details a randomized SVD algorithm that allows highly-scalable implementation of the method. In Section 4, we present a numerical experiment to illustrate the theoretical results established in Section 2 and test the computational performance of the randomized SVD approach to implement the optimal low-rank approximations. Our example consists of a high-dimensional atmospheric-transport source inversion problem using a large dataset of satellite observations. Finally, in Section 5 we investigate the links between the proposed optimal approximations and variational optimization algorithms used in current operational DA systems for NWP, and we propose a new Randomized Incremental Optimal Technique (RIOT) for 4D-Var based on our findings.

2 Theory

2.1 The Bayesian Problem

2.1.1 Finding the Maximum Likelihood

Here we shall review the Bayesian inversion approach to finding the maximum likelihood of a set of random variables, given some prior probability distribution functions (pdf) on these variables and on a set of physically-related observations, adopting the notations generally used in the numerical weather prediction community. Formally, the vector of observations, \mathbf{y} , is related to the so-called control vector \mathbf{x} through a forward model operator, H :

$$\mathbf{y} = H(\mathbf{x}), \quad (1)$$

where $\mathbf{x} \in E$, $\mathbf{y} \in F$, $H : E \rightarrow F$, and E and F are the control space (of dimension n) and the observation space (of dimension p), respectively.

Assuming Gaussian pdfs for the prior (\mathbf{x}^b) and the observations, with covariance error matrices \mathbf{B} and \mathbf{R} , respectively, the maximum likelihood can be obtained by minimizing the following cost function:

$$J(\mathbf{x}) = \frac{1}{2}(H(\mathbf{x}) - \mathbf{y})^T \mathbf{R}^{-1}(H(\mathbf{x}) - \mathbf{y}) + \frac{1}{2}(\mathbf{x} - \mathbf{x}^b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^b). \quad (2)$$

An analytical solution of (2) can be expressed as:

$$\mathbf{x}^a = \mathbf{x}^b + (\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{R}^{-1} (\mathbf{y} - H(\mathbf{x}^b)) \quad (3)$$

where \mathbf{H} is the Jacobian of the forward model. By applying the Sherman-Morrison-Woodbury formula to (3) (Sherman and Morrison, 1949), an alternative expression can be obtained:

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{B} \mathbf{H}^T (\mathbf{R} + \mathbf{H} \mathbf{B} \mathbf{H}^T)^{-1} (\mathbf{y} - H(\mathbf{x}^b)), \quad (4)$$

Equations (4) and (3) differ significantly in term of practical implementation. Eq. (4) requires forming and inverting the $(p \times p)$ matrix $\mathbf{R} + \mathbf{H} \mathbf{B} \mathbf{H}^T$, while in Eq. (3) the $(n \times n)$ matrix $\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}$ is inverted. The matrix $\mathbf{R} + \mathbf{H} \mathbf{B} \mathbf{H}^T$ is called the matrix of innovation statistics, and it plays an important role in DA methods.

For the common class of problems associated with a large control vector (e.g., $n > 10^6$) but a small number of observations ($p \ll n$), provided that a tangent linear (i.e., an implicit \mathbf{H}) and an adjoint (i.e., an implicit \mathbf{H}^T) models are available, it may be possible to explicitly form and invert the matrix of innovation statistics and to compute the maximum likelihood exactly using Eq. (4) (note that in this case \mathbf{B} would need to be defined implicitly as well). This approach is at the core of the representer method (Bennett, 2005), which is similar to the so-called Physical Space Assimilation System (PSAS). Note that in practice p adjoint and tangent-linear model integrations are required to extract the p columns of $\mathbf{H} \mathbf{B} \mathbf{H}^T$. Although parallel implementation is possible to compute those p columns, operational constraints (e.g., in NWP) and the limitation of computer resources may render this method impractical even for moderately large p (e.g., $p > 10^3$). In the case where n is small enough, the maximum likelihood solution can be obtained following a similar approach, but using (3) instead of (4). The formulation (3) is sometimes used in ensemble-based DA methods (e.g., EAKF) (Anderson, 2001), where a small number of perturbed trajectories is used to produce a sample estimate of $\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}$. If neither of the analytical formulations (4) and (3) can be directly used (e.g., if both n and p are very large), a variational optimization approach consisting of minimizing the cost function (2) is usually the method of choice, provided an adjoint model is available. However, solutions obtained from iterative minimization techniques are often only approximations to the maximum likelihood solution, since in practice the iteration is halted before full convergence is reached.

In the present study, we shall assume that n is very large ($n > 10^6$) and propose optimal approximations to the Bayesian solution whose practical implementations present good scalability properties. The optimality criteria considered will rely on the information content of the inversion, whose rigorous definition is the object of the following Section.

2.1.2 The Linear Case: Information Content and Incremental Formulation

In this study we shall assume that the forward model H is linear, so that $H = \mathbf{H}$. The non-linear case will be treated in Section 5, which investigates operational DA assimilation methods in NWP. Assuming linearity for the forward model allows us to rigorously define and compute useful quantities characterizing the information content of the inversion (Rodgers, 2000). With a linear

forward model, \mathbf{H} , the posterior distribution is Gaussian and the maximum likelihood is equal to the posterior mean. If the linear approximation of the forward model is valid in a neighborhood of the maximum-likelihood, the local posterior pdf is approximately Gaussian, in which case the notion of posterior error covariance becomes (locally) meaningful. Moreover, for a linear forward model, the cost function defined in (2) becomes quadratic, and the inverse of its Hessian at the minimum is equal to the posterior error covariance matrix, that is:

$$\mathbf{P}^a \equiv \overline{(\mathbf{x}^a - \mathbf{x}^t)(\mathbf{x}^a - \mathbf{x}^t)^T} = (\nabla^2 J)^{-1}(\mathbf{x}^a) = (\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1}, \quad (5)$$

where $\bar{\mathbf{x}}$ denotes the expectation of the random vector \mathbf{x} and \mathbf{x}^t represents the true state. Another useful formulation for \mathbf{P}^a can be obtained by applying the Sherman-Morrison-Woodbury formula to (5):

$$\mathbf{P}^a = \mathbf{B} - \mathbf{B} \mathbf{H}^T (\mathbf{H} \mathbf{B} \mathbf{H}^T + \mathbf{R})^{-1} \mathbf{H} \mathbf{B} \quad (6)$$

This formula expresses \mathbf{P}^a as a negative update of \mathbf{B} . The update term $(\mathbf{B} \mathbf{H}^T (\mathbf{H} \mathbf{B} \mathbf{H}^T + \mathbf{R})^{-1} \mathbf{H} \mathbf{B})$ can be interpreted as the posterior error reduction afforded by the observations. Another useful metric related to the information content of the problem is the DOFS, which quantifies the number of parameters independently constrained by the observations. It can be defined as the trace of the model resolution matrix (or averaging kernel) \mathbf{A} , which represents the sensitivity of the posterior mean to the true state (Rodgers, 2000):

$$\mathbf{A} \equiv \frac{\partial \mathbf{x}^a}{\partial \mathbf{x}^t} = \mathbf{Id} - \mathbf{P}^a \mathbf{B}^{-1} \quad (7)$$

$$\text{DOFs} = \text{Tr}(\mathbf{A}) \quad (8)$$

Finally, an additional useful formulation is to link the model resolution matrix to the posterior update term in (6):

$$\mathbf{A} = \mathbf{B} \mathbf{H}^T (\mathbf{H} \mathbf{B} \mathbf{H}^T + \mathbf{R})^{-1} \mathbf{H} \quad (9)$$

From (7) we clearly see that \mathbf{A} can be interpreted as a relative posterior error reduction.

Both $\mathbf{B} - \mathbf{P}^a$ and \mathbf{A} characterize the information content of the inversion (in an absolute and relative sense, respectively) and will be central to our analysis. Since the triplet $(\mathbf{x}^a, \mathbf{P}^a, \mathbf{A})$ fully characterizes the posterior pdf and the information content of the linear Bayesian problem, it shall be referred to as the solution of the Bayesian problem. It is worth noting that in our large-scale framework the matrices \mathbf{P}^a and \mathbf{A} cannot be computed directly nor represented explicitly in computer memory. Meaningful approximations of these quantities are therefore needed to properly interpret the statistical significance and the information content of the estimated posterior mean. Other applications include posterior sampling strategies (e.g., in cycling DA methods), where optimal and efficient approximations of the square-root for \mathbf{P}^a are required (see Section 5.2.2).

In the case of a linear forward model, \mathbf{H} , the posterior update formula (4) suggests a simplification of the problem by considering the increment $\delta \mathbf{x} \equiv \mathbf{x} - \mathbf{x}^b$ and innovation $\mathbf{d} \equiv \mathbf{y} - \mathbf{H} \mathbf{x}^b$ as the control and observation vector, respectively. The error statistics associated with the variables $\delta \mathbf{x}$ and \mathbf{d} are the same as those associated with \mathbf{x} and \mathbf{y} , respectively. Therefore, the previous equations defining the Bayesian solutions are unchanged when applying this change of variable. Note that in this incremental framework the prior is now the constant null vector ($\delta \mathbf{x}^b = 0$), whereas the true state is a random variable ($\delta \mathbf{x}^t = \mathbf{x}^t - \mathbf{x}^b$). In the rest of this paper, unless specified otherwise, the control vector \mathbf{x} will be identified with the increment $\delta \mathbf{x}$, and the observation vector \mathbf{y} replaced by the innovation vector \mathbf{d} .

2.1.3 Low-Rank Projections versus Low-Rank Approximations

When approximating the solution of a large-scale Bayesian problem, a fundamental distinction needs to be made between low-rank projections and low-rank approximations. A low-rank projection consists of restricting the Bayesian problem to a (small) subspace of the initial control space (by means of a projection operator). In this case a Bayesian problem of lower-rank is solved, and its solution can be considered to be an approximation of the initial problem in the sense that its posterior mean and posterior error covariance matrix converge to the true solutions as the reduced control space is (incrementally) increased. On the other hand, low-rank approximations of Bayesian problems belong to a more general class of methods that construct approximations of the posterior mean and posterior error covariance matrix of the initial high-dimensional problem, without the requirement of consistency between the approximated (low-rank) posterior mean and posterior error covariance (that is, they do not necessarily represent the posterior mean and corresponding posterior error covariance of a Bayesian problem). This distinction is important for interpretation as well as for applications of these methods. For instance, in a non-linear framework, a projection can be useful to define a low-rank version of a large-scale Bayesian problem to which MCMC sampling methods can be efficiently applied (e.g., Cui *et al.* (2014)). Other approximations of the same rank that do not correspond to a projected Bayesian problem may provide better estimates of the true solution, but would not be suitable for this application. In our study, we will first describe the formalism of low-rank projections and provide an optimal projection for the Bayesian problem that maximizes the information content (i.e., the DOFS) of the inversion (see Section 2.2). We will then explore the link between the solutions of this optimal projected problem and optimal low-rank approximations of the posterior mean and posterior error covariance matrix (Section 2.3).

2.2 Low-Rank Projections

2.2.1 General Formulation

One way to reduce the computational cost associated with solving a large-scale Bayesian problem is to project the problem onto a small subspace, $E' \subset E$, of dimension $k \ll n$. By construction, the projection restricts the posterior updates to the prior mean (\mathbf{x}^b) and to the prior error covariance matrix to the subspace E' , which effectively amounts to solving a problem of dimension k . The projection can be chosen so as to optimize some criteria, usually related to the information content of the inversion (e.g., maximum DOF or minimum posterior error covariance matrix for some norm). An important aspect of the projection is that it may induce an additional observational error if the observed subspace (i.e., the orthogonal of the kernel of \mathbf{H}) is not included in the range of the projector. This additional term is the so-called representativeness error. In the following we shall rigorously define the projected Bayesian problem and provide an analytical expression for the representativeness error.

Definition 2.1 (Projected Bayesian Problem). Let us consider a Bayesian problem defined by $\mathcal{B} \equiv (E, F, \mathbf{H}, \mathbf{B}, \mathbf{R})$ (using the definitions in Section 2.1.1), and a projection operator Π (i.e., $\Pi^2 = \Pi$). The projected problem associated with Π is the Bayesian problem $\mathcal{B}_\Pi \equiv (E_\Pi, F, \mathbf{H}_\Pi, \mathbf{B}_\Pi, \mathbf{R}_\Pi)$, where $E_\Pi = \{\Pi\mathbf{x}, \mathbf{x} \in E\}$, and \mathbf{H}_Π , \mathbf{B}_Π , and \mathbf{R}_Π are the forward model, prior and observation error covariance matrices, respectively, in some basis of E_Π and F .

The observational error covariance matrix (\mathbf{R}_Π) of the projected problem can be expressed as a function of \mathbf{B} and Π :

Proposition 2.2 (Representativeness Error). *The observational error covariance matrix \mathbf{R}_Π for the projected Bayesian problem $\mathcal{B}_\Pi = (E, F, \mathbf{H}\Pi, \mathbf{B}_\Pi, \mathbf{R}_\Pi)$ can be expressed as the sum of*

the observational error covariance for the original Bayesian problem, \mathbf{R} , and a representativeness error, as follows:

$$\mathbf{R}_\Pi = \mathbf{R} + \mathbf{H}(\mathbf{B} + \Pi\mathbf{B}\Pi^T - \mathbf{B}\Pi^T - \Pi\mathbf{B})\mathbf{H}^T \quad (10)$$

Proof. For the sake of clarity, below we distinguish between the control vector \mathbf{x} and its associated increment $\delta\mathbf{x} \equiv \mathbf{x} - \mathbf{x}^b$, and the observation vector \mathbf{y} and the corresponding *innovation* $\mathbf{d} \equiv \mathbf{y} - \mathbf{H}\mathbf{x}^b$. In the incremental framework, the observational error covariance matrix can be written:

$$\begin{aligned} \mathbf{R}_\Pi &\equiv \overline{(\mathbf{H}\Pi\delta\mathbf{x}^t - \mathbf{d})(\mathbf{H}\Pi\delta\mathbf{x}^t - \mathbf{d})^T} \\ &= \overline{(\mathbf{H}\Pi\mathbf{x}^t - \mathbf{H}\Pi\mathbf{x}^b + \mathbf{H}\mathbf{x}^b - \mathbf{H}\mathbf{x}^t + \epsilon)(\mathbf{H}\Pi\mathbf{x}^t - \mathbf{H}\Pi\mathbf{x}^b + \mathbf{H}\mathbf{x}^b - \mathbf{H}\mathbf{x}^t + \epsilon)^T} \end{aligned}$$

Using the independence assumption between the errors in the observations and in the prior, one obtains:

$$\mathbf{R}_\Pi = \mathbf{R} + \mathbf{H}(\mathbf{B} + \Pi\mathbf{B}\Pi^T - \mathbf{B}\Pi^T - \Pi\mathbf{B})\mathbf{H}^T$$

□

Our goal is to find a projection that maximizes the DOFS or minimizes the posterior error covariance matrix of the Bayesian problem \mathcal{B}_Π , in some sense to be defined thereafter. In the following we describe a two-step approach to the optimal projection problem, wherein an appropriate decomposition is used to construct a class of projectors in which the optimal solutions must lie. This restriction to a particular class of projectors allows us to greatly simplify the problem, as we show in Section 2.2.4. In addition, the two-step method yields some interesting theoretical interpretations, and can be related to previous Bayesian dimension reduction approaches, as described in Sections 2.2.2 and 2.2.3.

2.2.2 A Two-Step Approach

The idea behind the projection approach is to solve a Bayesian problem of smaller dimension (k) than the original large-scale problem (i.e., one has $k \ll n$), allowing fast (sometimes analytical) computation of its solution. In this Section, a factorization of rank- k projectors is proposed to construct a Bayesian problem of dimension k from which the solutions of the projected problem \mathcal{B}_Π in the canonical basis of E can be derived with simple transformations. Let us recall that any projector is defined by its null space and its range, and can be written:

$$\Pi = \mathbf{I}(\mathbf{O}^T\mathbf{I})^{-1}\mathbf{O}^T, \quad (11)$$

where \mathbf{O} is a matrix whose columns form an orthonormal basis for the orthogonal of the null space of Π , and \mathbf{I} is a matrix whose columns span the range of Π . In other words, \mathbf{I} defines the subspace (of dimension k) onto which the Bayesian problem is projected, while \mathbf{O} defines the direction of the projection. Another form for (11) can be derived as follows:

Proposition 2.3 (Factorization of a Projector). *Any linear operator Π is a projector of rank k if and only if it can be written as the product of two rank- k matrices, one of which is the left inverse of the other, i.e.:*

$$\Pi = \mathbf{\Gamma}^*\mathbf{\Gamma}, \quad (12)$$

where $\mathbf{\Gamma}^*$ and $\mathbf{\Gamma}$ two matrices of dimension $(n \times k)$ and $(k \times n)$, respectively, with maximum rank and:

$$\mathbf{\Gamma}\mathbf{\Gamma}^* = \mathbf{Id}_k \quad (13)$$

Proof. Any projector Π of rank k can be written $\Pi = \mathbf{I}(\mathbf{O}^T \mathbf{I})^{-1} \mathbf{O}^T$, where \mathbf{I} (the range of Π) and \mathbf{O} (the orthogonal of the null space of Π) are two matrices of rank k and dimension $(n \times k)$. Defining $\Gamma = \mathbf{O}^T$ and $\Gamma^* = \mathbf{I}(\mathbf{O}^T \mathbf{I})^{-1}$, we obtain $\Pi = \Gamma^* \Gamma$, with Γ and Γ^* of dimension $(n \times k)$ and $(k \times n)$, respectively, and $\Gamma \Gamma^* = \mathbf{Id}_k$. Now let us consider a linear operator $\Pi = \Gamma^* \Gamma$, with $\Gamma \Gamma^* = \mathbf{Id}_k$. One has also $\Pi = \Gamma^* (\Gamma \Gamma^*)^{-1} \Gamma$. Defining $\mathbf{O} = \Gamma^T$ and $\mathbf{I} = \Gamma^*$, one sees that Π has the general form of a projector. \square

Remarks 2.1. In practice, Γ and Γ^* can be derived from the range and the null space of the projection Π (see previous proof). Note that the decomposition $\Pi = \Gamma^* \Gamma$ is not unique. Indeed, $\Gamma = (\mathbf{O} \mathbf{P})^T$ and $\Gamma^* = \mathbf{I} \mathbf{Q} ((\mathbf{O} \mathbf{P})^T \mathbf{I} \mathbf{Q})^{-1}$ verify (12) and (13) for all orthogonal matrices \mathbf{P} and \mathbf{Q} (i.e., verifying $\mathbf{P}^T \mathbf{P} = \mathbf{Q}^T \mathbf{Q} = \mathbf{Id}$). This simply formalizes the fact that a projection is only defined by its null space and range, and is therefore basis-invariant.

The previous decomposition is now used to define a reduced Bayesian problem of dimension k in order to be able to express the solution of the projected problem in the initial control space E .

Definition 2.4 (Reduced Bayesian Problem). Let us consider a projected Bayesian problem defined by $\mathcal{B}_\Pi \equiv (E_\Pi, F, \mathbf{H}_\Pi, \mathbf{B}_\Pi, \mathbf{R}_\Pi)$. The *reduced Bayesian problem* associated with \mathcal{B}_Π in the basis defined by the columns of Γ^* is the problem $\mathcal{B}_\omega = (E_\omega, F, \mathbf{H} \Gamma^*, \Gamma \mathbf{B} \Gamma^T, \mathbf{R}_\Pi)$, where $E_\omega = \{\Gamma \mathbf{x}, \mathbf{x} \in E\}$ and $\Pi = \Gamma^* \Gamma$ is a rank- k factorization of the projector Π , as defined in Prop. 2.3.

The couple (Γ, Γ^*) characterizes the correspondence between the reduced control space E_ω , on which the reduced problem \mathcal{B}_ω is defined, and the initial control space E . Note that the k columns of Γ^* form a basis for the subspace of E on which the Bayesian problem is projected. Therefore, the vector $\Gamma \mathbf{x}$ represents the coordinates of \mathbf{x} in the basis defined by Γ^* . Using the posterior solution of the reduced Bayesian problem \mathcal{B}_ω , one can provide an analytical solution for the projected problem \mathcal{B}_Π in the initial space E :

Proposition 2.5 (Posterior Solution for a General Projection). *Let us consider a projector $\Pi = \Gamma^* \Gamma$, factored according to Prop. 2.3. We define the associated projected and reduced Bayesian problems, $\mathcal{B}_\Pi \equiv (E_\Pi, F, \mathbf{H}_\Pi, \mathbf{B}_\Pi, \mathbf{R}_\Pi)$ and $\mathcal{B}_\omega = (E_\omega, F, \mathbf{H} \Gamma^*, \Gamma \mathbf{B} \Gamma^T, \mathbf{R}_\Pi)$, respectively. One has:*

$$\mathbf{x}_\Pi^a = \Gamma^* \mathbf{x}_\omega^a \quad (14)$$

$$\mathbf{P}_\Pi^a = \Gamma^* \mathbf{P}_\omega^a \Gamma^{*T} \quad (15)$$

$$\mathbf{A}_\Pi = \Gamma^* \mathbf{A}_\omega \Gamma, \quad (16)$$

where \mathbf{x}_Π^a and \mathbf{x}_ω^a are the posterior mean of \mathcal{B}_Π and \mathcal{B}_ω , respectively, \mathbf{P}_Π^a and \mathbf{P}_ω^a are the posterior error covariance matrices of \mathcal{B}_Π and \mathcal{B}_ω , respectively, and \mathbf{A}_Π and \mathbf{A}_ω are the model resolution matrices of \mathcal{B}_Π and \mathcal{B}_ω , respectively. More precisely:

$$\mathbf{x}_\Pi^a = \Pi \mathbf{B} \Pi^T \mathbf{H}^T (\Pi \mathbf{B} \Pi^T \mathbf{H}^T + \mathbf{R}_\Pi)^{-1} \mathbf{d} \quad (17)$$

$$\mathbf{P}_\Pi^a = \Pi \mathbf{B} \Pi^T - \Pi \mathbf{B} \Pi^T \mathbf{H}^T (\Pi \mathbf{B} \Pi^T \mathbf{H}^T + \mathbf{R}_\Pi)^{-1} \Pi \mathbf{B} \Pi^T \quad (18)$$

$$\mathbf{A}_\Pi = \Pi \mathbf{B} \Pi^T \mathbf{H}^T (\Pi \mathbf{B} \Pi^T \mathbf{H}^T + \mathbf{R}_\Pi)^{-1} \Pi \quad (19)$$

Or equivalently:

$$\mathbf{x}_\Pi^a = \Gamma^* \left[(\Gamma \mathbf{B} \Gamma^T)^{-1} + \Gamma^{*T} \mathbf{H}^T \mathbf{R}_\Pi^{-1} \mathbf{H} \Gamma^* \right]^{-1} \Gamma^{*T} \mathbf{H}^T \mathbf{R}_\Pi^{-1} \mathbf{d} \quad (20)$$

$$\mathbf{P}_\Pi^a = \Gamma^* \left[(\Gamma \mathbf{B} \Gamma^T)^{-1} + \Gamma^{*T} \mathbf{H}^T \mathbf{R}_\Pi^{-1} \mathbf{H} \Gamma^* \right]^{-1} \Gamma^{*T} \quad (21)$$

$$\mathbf{A}_\Pi = \Pi - \Gamma^* \left[(\Gamma \mathbf{B} \Gamma^T)^{-1} + \Gamma^{*T} \mathbf{H}^T \mathbf{R}_\Pi^{-1} \mathbf{H} \Gamma^* \right]^{-1} (\Gamma \mathbf{B} \Gamma^T)^{-1} \Gamma \quad (22)$$

Here the solution $(\mathbf{x}_\Pi^a, \mathbf{P}_\Pi^a, \mathbf{A}_\Pi)$ is expressed in the canonical basis of the initial control space E .

Proof. Let us define the two vector spaces $E_\Pi = \{\Pi\mathbf{x}, \mathbf{x} \in E\}$ and $E_\omega = \{\Gamma\mathbf{x}, \mathbf{x} \in E\}$. One can easily verify that the following application defines an isomorphism between E_Π and E_ω :

$$\begin{cases} g : E_\Pi \rightarrow E_\omega \\ \mathbf{x}_\Pi \mapsto \mathbf{x}_\omega = \Gamma\mathbf{x}_\Pi \\ g^{-1} : E_\omega \rightarrow E_\Pi \\ \mathbf{x}_\omega \mapsto \mathbf{x}_\Pi = \Gamma^*\mathbf{x}_\omega \end{cases}$$

With this definition, g associates any vector of E_Π expressed in the canonical basis of E to its coordinates in the basis formed by the columns of Γ^* , while g^{-1} associates any vector of E_Π expressed in the basis defined by Γ^* to its coordinates in the canonical basis of E . The prior error covariance matrix \mathbf{B}_Π in the basis defined by Γ^* is simply $\overline{(\Gamma\Pi(\mathbf{x}^t - \mathbf{x}^b)(\Gamma\Pi(\mathbf{x}^t - \mathbf{x}^b))^T} = \overline{(\Gamma(\mathbf{x}^t - \mathbf{x}^b)(\Gamma(\mathbf{x}^t - \mathbf{x}^b))^T} = \Gamma\mathbf{B}\Gamma^T$. Likewise, the forward model \mathbf{H}_Π expressed in the basis defined by Γ^* is simply $\mathbf{H}\Gamma^*$, since: $\forall \mathbf{x}_\Pi \in E_\Pi, \mathbf{H}_\Pi\mathbf{x}_\Pi = \mathbf{H}\Pi\mathbf{x} = \mathbf{H}\Gamma^*\Gamma\Pi\mathbf{x} = \mathbf{H}\Gamma^*\Gamma\mathbf{x} = \mathbf{H}\Gamma^*\mathbf{x}_\omega$. Therefore, the Bayesian problems $\mathcal{B}_\Pi \equiv (E_\Pi, F, \mathbf{H}_\Pi, \mathbf{B}_\Pi, \mathbf{R}_\Pi)$ and $\mathcal{B}_\omega = (E_\omega, F, \mathbf{H}\Gamma^*, \Gamma\mathbf{B}\Gamma^T, \mathbf{R}_\Pi)$ are strictly equivalent (\mathcal{B}_ω is \mathcal{B}_Π expressed in the particular basis defined by Γ^*). Noting $\mathbf{x}_\omega^a, \mathbf{P}_\omega^a$ and \mathbf{A}_ω the posterior mean, posterior error covariance and model resolution matrix, respectively, of the Bayesian problem \mathcal{B}_ω , one can directly obtain Eq. (14) and (15) by applying g^{-1} . For Eq. (16), we note that the model resolution matrix of the reduced problem \mathcal{B}_Π in the canonical basis must verify: $\forall \mathbf{x} \in E, \mathbf{A}_\Pi\Pi\mathbf{x} = \Gamma^*\mathbf{A}_\omega\mathbf{x}_\omega$. Using $\mathbf{x}_\omega = \Gamma\Pi\mathbf{x}$ in the right-hand side, one obtains $\mathbf{A}_\Pi\Pi\mathbf{x} = \Gamma^*\mathbf{A}_\omega\Gamma\Pi\mathbf{x}$, which by identification gives Eq. (16). Formulas (17)-(19) and (20)-(22) are then obtained by replacing $(\mathbf{x}_\omega^a, \mathbf{P}_\omega^a, \mathbf{A}_\omega)$ in (14)-(16) using Eq. (4), (6), (9), and Eq. (3), (5), (7), respectively. \square

Remarks 2.2. As long as the dimension p of the observation space F allows for explicit construction and inversion of $(p \times p)$ covariance matrices in F , formulas (17)-(19) or (20)-(22) can be used to compute the posterior mean and extract any column of the posterior error covariance or model resolution matrices for the projected problem. On the other hand, when p is large, due to the presence of the representativeness error \mathbf{R}_Π and the necessity to form the $(p \times p)$ matrix $\mathbf{H}(\mathbf{B} + \Pi\mathbf{B}\Pi^T - \mathbf{B}\Pi^T - \Pi\mathbf{B})\mathbf{H}^T$, it may not be practical to use either of the two formulations (17)-(19) or (20)-(22). Interestingly, as we will show in Section 2.2.4, one can define an optimal projection which has the remarkable property that it effectively avoids the need to know and evaluate the representativeness errors covariance matrix to compute the posterior solution.

Aggregations and Projections In the multi-scale formalism presented by Bocquet *et al.* (2011), Γ and Γ^* are called the *aggregation* and *prolongation* operators, respectively, and $\Gamma\Gamma^* = \mathbf{Id}_k$ is an imposed stability condition. In their study the operator Γ consists of a weighted average of model grid cell parameters (e.g., atmospheric fluxes) and the objective is to solve an aggregated version of the initial Bayesian problem of smaller dimension, which corresponds to our reduced problem \mathcal{B}_ω . Although Prop. 2.5 shows that the formulations for the aggregated problem \mathcal{B}_ω and the projected problem \mathcal{B}_Π are theoretically equivalent, their interpretations are quite different. In the projection framework, the analysis is centered on choosing a subspace of E , E_Π , on which the Bayesian problem is solved, i.e., on the choice of Γ^* , whose columns represent a basis of E_Π (e.g., Turner and Jacob (2015)). On the other hand, in the aggregation framework, the problem focuses on the choice of an average operator to define an aggregated problem, i.e., on Γ . Likewise, in the projection framework the posterior solution is analyzed in E_Π , while in the aggregation framework the posterior solution for the aggregated control vector in E_ω is the meaningful quantity to interpret. Note that, from Eq. (16), the DOFS of the aggregated problem is the same as the DOFS of the associated projected problem expressed in the canonical basis of E , since $\text{Tr}(\mathbf{A}_\Pi) = \text{Tr}(\Gamma^*\mathbf{A}_\omega\Gamma) = \text{Tr}(\mathbf{A}_\omega\Gamma\Gamma^*) = \text{Tr}(\mathbf{A}_\omega)$. In our analysis Γ is a general $(k \times n)$ operator,

therefore we shall refer to it as a *reduction* operator (instead of an aggregation operator) and refer to $\mathbf{\Gamma}^*$ as a prolongation operator for the sake of consistency with the formalism of Bocquet *et al.* (2011).

2.2.3 A Generalized Change of Variable for Linear Bayesian Problems

We now turn to the problem of optimizing the choice of the projection $\mathbf{\Pi}$. The factorization of the projection described in the previous section will be used to construct our optimal solution in two steps. The first step consists, for a given reduction operator $\mathbf{\Gamma}$, of finding a prolongation operator $\mathbf{\Gamma}^*$ that minimizes the representativeness error $\mathbf{R}_{\mathbf{\Pi}}$ of the projected problem. The following Theorem provides such an optimal prolongation operator $\mathbf{\Gamma}^*$ as a function of $\mathbf{\Gamma}$:

Theorem 2.6 (Optimal Prolongation). *For any reduction operator $\mathbf{\Gamma}$, there exists a prolongation operator $\mathbf{\Gamma}_{opt}^*$ such that the representativeness error is minimum w.r.t. the Löwner partial ordering. More specifically, one has:*

$$\forall \mathbf{\Gamma} \in \mathcal{M}_{k,n}(\mathbb{R}), \exists \mathbf{\Gamma}_{opt}^* \in \mathcal{M}_{n,k}(\mathbb{R}) \mid \forall \mathbf{\Gamma}^* \in \mathcal{M}_{n,k} : \mathbf{R}_{\mathbf{\Pi}_{opt}} \leq \mathbf{R}_{\mathbf{\Pi}}, \quad (23)$$

where $\mathbf{\Pi} = \mathbf{\Gamma}^* \mathbf{\Gamma}$, $\mathbf{\Pi}_{opt} = \mathbf{\Gamma}_{opt}^* \mathbf{\Gamma}$, $\mathcal{M}_{m,n}$ represents the space of $(m \times n)$ real matrices, and the symbol \leq denotes the Löwner partial ordering within the set of real positive definite matrices. Moreover, one has:

$$\mathbf{\Gamma}_{opt}^* = \mathbf{B} \mathbf{\Gamma}^T (\mathbf{\Gamma} \mathbf{B} \mathbf{\Gamma}^T)^{-1} \quad (24)$$

Proof. Let us rewrite the observational error covariance for the projected problem using the decomposition (2.3):

$$\mathbf{R}_{\mathbf{\Pi}} = \mathbf{R} + \mathbf{H}(\mathbf{B} + \mathbf{\Gamma}^* \mathbf{\Gamma} \mathbf{B} \mathbf{\Gamma}^T \mathbf{\Gamma}^{*T} - \mathbf{B} \mathbf{\Gamma}^T \mathbf{\Gamma}^{*T} - \mathbf{\Gamma}^* \mathbf{\Gamma} \mathbf{B}) \mathbf{H}^T \quad (25)$$

From Lemma (118), it is clear that minimizing $\mathbf{R}_{\mathbf{\Pi}}$ is equivalent to minimizing the matrix $\Delta \mathbf{B} = \mathbf{B} + \mathbf{\Gamma}^* \mathbf{\Gamma} \mathbf{B} \mathbf{\Gamma}^T \mathbf{\Gamma}^{*T} - \mathbf{B} \mathbf{\Gamma}^T \mathbf{\Gamma}^{*T} - \mathbf{\Gamma}^* \mathbf{\Gamma} \mathbf{B}$. Fixing $\mathbf{\Gamma}$, we note that the solution (best prolongation $\mathbf{\Gamma}_{opt}^*$) to this minimization problem is also the Best Linear Unbiased Estimator (BLUE) of the following problem:

$$\begin{aligned} & \text{Arg min}_{\mathbf{\Gamma}^*} \overline{\text{Tr}(\mathbf{x} - \mathbf{x}^t)(\mathbf{x} - \mathbf{x}^t)^T}, \\ & \text{with } \begin{cases} \mathbf{x} = \mathbf{x}_b + \mathbf{\Gamma}^*(\mathbf{y} - \mathbf{\Gamma} \mathbf{x}_b) \\ \mathbf{y} = \mathbf{\Gamma} \mathbf{x}^t \\ \overline{(\mathbf{x}^b - \mathbf{x}^t)(\mathbf{x}^b - \mathbf{x}^t)^T} = \mathbf{B} \end{cases} \end{aligned}$$

The BLUE solution to this problem, and therefore the optimal prolongation operator $\mathbf{\Gamma}^*$, is given by $\mathbf{\Gamma}_{opt}^* = \mathbf{B} \mathbf{\Gamma}^T (\mathbf{\Gamma} \mathbf{B} \mathbf{\Gamma}^T)^{-1}$. The posterior error covariance matrix of the BLUE analysis is precisely $\Delta \mathbf{B}$, and it is minimum in the sense of the Löwner partial ordering among all linear estimator (i.e., among all prolongation operators) (e.g., Isotalo *et al.* (2008)), which proves (23). \square

Remarks 2.3. The optimal prolongation $\mathbf{\Gamma}_{opt}^*$ was first proposed by Bocquet *et al.* (2011), where it was derived from a Bayesian perspective exploiting the prior information. In our approach this result is obtained simply by minimizing the representativeness error.

Theorem 2.6 is a strong optimality result, since the optimality w.r.t. the Löwner partial ordering implies that the representativeness error is minimum in any direction of the observation space. This leads to the following important optimality result:

Corollary 2.7. *For any reduction operator Γ , the prolongation operator $\Gamma_{opt}^* = \mathbf{B}\Gamma^T(\Gamma\mathbf{B}\Gamma^T)^{-1}$ minimizes the Fisher measurement information matrix w.r.t. the Löwner partial ordering, i.e.:*

$$\forall \Gamma \in \mathcal{M}_{k,n}(\mathbb{R}), \exists \Gamma_{opt}^* \in \mathcal{M}_{n,k}(\mathbb{R}) \mid \forall \Gamma^* \in \mathcal{M}_{n,k} : \mathbf{H}^T \mathbf{R}_{\Pi_{opt}}^{-1} \mathbf{H} \leq \mathbf{H}^T \mathbf{R}_{\Pi}^{-1} \mathbf{H}, \quad (26)$$

where $\Pi = \Gamma^* \Gamma$ and $\Pi_{opt} = \Gamma_{opt}^* \Gamma$.

Proof. This follows by applying consecutively Lemmas (117) and (118) to (23). \square

Theorem 2.6 and the optimal prolongation (24) yield different interpretations depending on the application. In the context of aggregation (see 2.2.2), once an aggregation operator Γ (e.g., a weighted average of model grid-cells) has been chosen, the optimal prolongation (24) should be constructed and used together with the (reduced) forward model $\mathbf{H}\Gamma^*$. On the other hand, in the context of a low-rank projection, once a subspace for the range of the projector has been chosen, Eq. (24) imposes an optimal direction for the projection. More precisely, if the columns of the matrix \mathbf{I} represent a basis for the range of the projector Π , then an optimal direction is defined by $\mathbf{D} = \mathbf{I} - \mathbf{B}^{-1}\mathbf{I}$ (i.e., $\mathbf{O} = \mathbf{B}^{-1}\mathbf{I}$ in (11)). As an example of application of those concepts, we note that, in Turner and Jacob (2015), the computation of the Gaussian Mixture Model (GMM) basis defining the range of the projection is performed simultaneously with the computation of the direction of the projection, or equivalently, the operators Γ^* and Γ are constructed all at once. The fact that the resulting projection does not belong to the class of optimal projection defined in Prop. 2.11 is revealed by the presence of suboptimal features, such as a non-trivial minimum in the total posterior error variance for a rank $k < n$ (see interactive discussion of Turner and Jacob (2015)). Based on our results, one approach to improve the method proposed by Turner and Jacob (2015) would be to use the computed Gaussian Mixture Model (GMM) basis as range for the projection, but to replace their reduction operator (i.e., their weight matrix \mathbf{W}) by one that corresponds to an optimal direction for the projection.

The following Proposition provides another useful optimality result for interpreting the optimal prolongation Γ_{opt}^* :

Proposition 2.8. *For any given reduction operator Γ , the associated optimal prolongation operator Γ_{opt}^* defines a rank- k projection Π_{opt} that minimizes, over all rank- k projections, the Frobenius distance between any square-root of the prior error covariance and its projection, i.e.:*

$$\|\mathbf{L} - \Pi_{opt}\mathbf{L}\|_F = \min_{\Pi \in \mathcal{P}} \|\mathbf{L} - \Pi\mathbf{L}\|_F, \quad (27)$$

where $\Pi_{opt} = \Gamma_{opt}^* \Gamma$, $\mathcal{P} = \{\Pi \in \mathcal{M}_n \mid \Pi^2 = \Pi, \text{rank}(\Pi) = k\}$ and $\mathbf{B} = \mathbf{L}\mathbf{L}^T$

Proof. The proof is simply obtained by choosing a square-root \mathbf{L} of \mathbf{B} (i.e., $\mathbf{B} = \mathbf{L}\mathbf{L}^T$ and noting that:

$$\begin{aligned} \Delta\mathbf{B} &= \mathbf{B} + \Pi\mathbf{B}\Pi^T - \mathbf{B}\Pi^T - \Pi\mathbf{B} \\ &= (\mathbf{L} - \Pi\mathbf{L})(\mathbf{L} - \Pi\mathbf{L})^T \end{aligned}$$

Since $\Delta\mathbf{B}$ is minimum for the Löwner partial ordering, one has in particular:

$$\begin{aligned} \text{Tr} [(\mathbf{L} - \Pi_{opt}\mathbf{L})(\mathbf{L} - \Pi_{opt}\mathbf{L})^T] &= \min_{\Pi} \text{Tr} [(\mathbf{L} - \Pi\mathbf{L})(\mathbf{L} - \Pi\mathbf{L})^T] \\ &\iff \\ \|\mathbf{L} - \Pi_{opt}\mathbf{L}\|_F &= \min_{\Pi} \|\mathbf{L} - \Pi\mathbf{L}\|_F \end{aligned}$$

\square

Finally, a simple interpretation for the optimal couple $(\mathbf{\Gamma}, \mathbf{\Gamma}_{opt}^*)$ is possible based on the following results:

Proposition 2.9 (Posterior Solution of the Reduced Bayesian Problem). *Let us define a Bayesian problem $\mathcal{B} = (E, F, \mathbf{H}, \mathbf{B}, \mathbf{R})$. Let us consider a reduction $\mathbf{\Gamma}$ and its optimal prolongation $\mathbf{\Gamma}_{opt}^*$, and $\mathcal{B}_\omega = (E_\omega, F, \mathbf{H}\mathbf{\Gamma}_{opt}^*, \mathbf{\Gamma}^T \mathbf{B} \mathbf{\Gamma}, \mathbf{R}_{\Pi_{opt}})$ the associated reduced Bayesian problem. One has:*

$$\mathbf{x}_\omega^a = \mathbf{\Gamma} \mathbf{x}^a \quad (28)$$

$$\mathbf{P}_\omega^a = \mathbf{\Gamma} \mathbf{P}^a \mathbf{\Gamma}^T \quad (29)$$

$$\mathbf{A}_\omega = \mathbf{\Gamma} \mathbf{A} \mathbf{\Gamma}_{opt}^*, \quad (30)$$

where \mathbf{x}^a , \mathbf{P}^a and \mathbf{A} are the posterior mean, posterior error covariance and model resolution matrix (respectively) of \mathcal{B} , and \mathbf{x}_ω^a , \mathbf{P}_ω^a and \mathbf{A}_ω are the posterior mean, posterior error covariance and model resolution matrix (respectively) of \mathcal{B}_ω .

Proof. Formulas (28)-(30) are obtained by using the optimality properties $\mathbf{\Pi}_{opt} \mathbf{B} \mathbf{\Pi}_{opt}^T = \mathbf{\Pi}_{opt} \mathbf{B} = \mathbf{B} \mathbf{\Pi}_{opt}^T$ and the (resulting) invariance of the innovation statistics $(\mathbf{H} \mathbf{B} \mathbf{H}^T + \mathbf{R})$ in formulas (17)-(19). \square

Similar formulas can be obtained for the solution of the projected Bayesian problem in the canonical basis of E :

Corollary 2.10 (Posterior Solution of the Projected Bayesian Problem). *Let us define a Bayesian problem $\mathcal{B} = (E, F, \mathbf{H}, \mathbf{B}, \mathbf{R})$. Let us consider a reduction $\mathbf{\Gamma}$ and its optimal prolongation $\mathbf{\Gamma}_{opt}^*$, and the associated projector $\mathbf{\Pi}_{opt} = \mathbf{\Gamma}_{opt}^* \mathbf{\Gamma}$. One has:*

$$\mathbf{x}_{\Pi_{opt}}^a = \mathbf{\Pi}_{opt} \mathbf{x}^a \quad (31)$$

$$\mathbf{P}_{\Pi_{opt}}^a = \mathbf{\Pi}_{opt} \mathbf{P}^a \mathbf{\Pi}_{opt}^T \quad (32)$$

$$\mathbf{A}_{\Pi_{opt}} = \mathbf{\Pi}_{opt} \mathbf{A} \mathbf{\Pi}_{opt}, \quad (33)$$

where \mathbf{x}^a , \mathbf{P}^a and \mathbf{A} are the posterior mean, posterior error covariance and model resolution matrix (respectively) of \mathcal{B} , and $\mathbf{x}_{\Pi_{opt}}^a$, $\mathbf{P}_{\Pi_{opt}}^a$ and $\mathbf{A}_{\Pi_{opt}}$ are the posterior mean, posterior error covariance and model resolution matrix (respectively) of the projected Bayesian problem $\mathcal{B}_{\Pi_{opt}} \equiv (E_{\Pi_{opt}}, F, \mathbf{H}_{\Pi_{opt}}, \mathbf{B}_{\Pi_{opt}}, \mathbf{R}_{\Pi_{opt}})$ in the canonical basis of E .

Proof. Formulas (31)-(33) are obtained simply by applying (14)-(16) to (28)-(30). \square

In other words, if the projector is of the form $\mathbf{\Pi}_{opt} = \mathbf{B} \mathbf{\Gamma}^T (\mathbf{\Gamma}^T \mathbf{B} \mathbf{\Gamma})^{-1} \mathbf{\Gamma}$, the solution of the projected Bayesian problem is simply the projection of the solution of the initial Bayesian problem. From (2.9), it is clear that the couple $(\mathbf{\Gamma}, \mathbf{\Gamma}_{opt}^*)$ can be interpreted as a generalized change of variable for the solutions of linear Bayesian problems, where the transformation $\mathbf{\Gamma}$ can be non-invertible and $\mathbf{\Gamma}_{opt}^*$ defines a right inverse for $\mathbf{\Gamma}$. It is straightforward to verify that in the case where $\mathbf{\Gamma}$ is invertible $\mathbf{\Gamma}_{opt}^* = \mathbf{\Gamma}^{-1}$.

Remarks 2.4. It is interesting to note that the posterior solution of the projected Bayesian problem $\mathcal{B}_{\Pi_{opt}} \equiv (E_{\Pi_{opt}}, F, \mathbf{H}_{\Pi_{opt}}, \mathbf{B}_{\Pi_{opt}}, \mathbf{R}_{\Pi_{opt}})$ is also the posterior solution of the Bayesian problem $\mathcal{B}_{\mathbf{H}\mathbf{\Pi}_{opt}} \equiv (E, F, \mathbf{H}\mathbf{\Pi}_{opt}, \mathbf{B}, \mathbf{R}_{\Pi_{opt}})$, as one can see by considering an optimal prolongation in formulas (17)-(19), and using $\mathbf{H}\mathbf{\Pi}\mathbf{\Pi}^T \mathbf{H}^T + \mathbf{R}_{\Pi} = \mathbf{H} \mathbf{B} \mathbf{H}^T + \mathbf{R}$ and $\mathbf{\Pi} \mathbf{B} \mathbf{\Pi}^T = \mathbf{\Pi} \mathbf{B} = \mathbf{B} \mathbf{\Pi}^T$. The Bayesian problem $\mathcal{B}_{\mathbf{H}\mathbf{\Pi}_{opt}}$ has similar prior error covariance as the original problem \mathcal{B} , but corresponds to a forward model for which the modes are filtered by the projection $\mathbf{\Pi}$. Note that $\mathcal{B}_{\Pi_{opt}}$ and $\mathcal{B}_{\mathbf{H}\mathbf{\Pi}_{opt}}$ do not in general define the same Bayesian problem. The equality between the posterior solutions of those two problems only holds when an optimal prolongation is used.

2.2.4 An Optimal Projection

The results established in Section 2.2.3, and in particular Theorem 2.6, can be used to simplify the problem of defining an optimal low-rank projection for the Bayesian problem $\mathcal{B}_\Pi \equiv (E_\Pi, F, \mathbf{H}_\Pi, \mathbf{B}_\Pi, \mathbf{R}_\Pi)$. Indeed, one can restrict our search to the class of projections associated with optimal prolongations, using (24). We first note that this class of projections can be redefined in a simpler form:

Proposition 2.11 (Canonical Form of Projections). *Let us define the class of projections of rank k associated with optimal prolongations $\mathcal{P}_{opt} \equiv \{\mathbf{B}\Gamma^T(\Gamma\mathbf{B}\Gamma^T)^{-1}\Gamma, \Gamma \in \mathcal{M}_{k,n}\}$. One has:*

$$\mathcal{P}_{opt} = \{\mathbf{B}^{1/2}\mathbf{U}\mathbf{U}^T\mathbf{B}^{-1/2}, \mathbf{U}^T\mathbf{U} = \mathbf{Id}_k \text{ and } \mathbf{U} \in \mathcal{M}_{k,n}\} \quad (34)$$

Proof. To prove that $\{\mathbf{B}\Gamma^T(\Gamma\mathbf{B}\Gamma^T)^{-1}\Gamma, \Gamma \in \mathcal{M}_{k,n}\} \subset \{\mathbf{B}^{1/2}\mathbf{U}\mathbf{U}^T\mathbf{B}^{-1/2}, \mathbf{U}^T\mathbf{U} = \mathbf{Id}_k\}$, we define $\mathbf{U} = \mathbf{B}^{1/2}\Gamma^T(\Gamma\mathbf{B}\Gamma^T)^{-1/2}$, and note that $\mathbf{B}^{1/2}\mathbf{U}\mathbf{U}^T\mathbf{B}^{-1/2} = \mathbf{B}\Gamma^T(\Gamma\mathbf{B}\Gamma^T)^{-1}\Gamma$ and $\mathbf{U}\mathbf{U}^T = \mathbf{Id}_k$. To prove that $\{\mathbf{B}^{1/2}\mathbf{U}\mathbf{U}^T\mathbf{B}^{-1/2}, \mathbf{U}^T\mathbf{U} = \mathbf{Id}_k\} \subset \{\mathbf{B}\Gamma^T(\Gamma\mathbf{B}\Gamma^T)^{-1}\Gamma, \Gamma \in \mathcal{M}_{k,n}\}$, we define $\Gamma = \mathbf{U}^T\mathbf{B}^{1/2}$, which verifies $\mathbf{B}\Gamma^T(\Gamma\mathbf{B}\Gamma^T)^{-1}\Gamma = \mathbf{B}^{1/2}\mathbf{U}\mathbf{U}^T\mathbf{B}^{-1/2}$. \square

Remarks 2.5. Prop. 2.11 allows a simple (statistical) interpretation for the optimal projections, that is, they correspond to a whitening transformation ($\mathbf{B}^{-1/2}$) followed by a orthogonal projection ($\mathbf{U}\mathbf{U}^T$) onto a rank- k subspace of E , and a coloring transformation ($\mathbf{B}^{1/2}$) that recovers the prior error covariances.

We can now state one of the main results of this paper, which provides an optimal low-rank projection for the linear Bayesian problem. The following Lemma provides a basis of eigenvectors for the model resolution matrix, which, together with the canonical form of Prop. 2.11, allows for construction of a projected Bayesian problem $\mathcal{B}_{\Pi_{opt}}$ with maximum DOFS. Note that in this paper eigenvectors shall always be presented in descending order of their corresponding eigenvalues.

Lemma 2.12 (Diagonalization of the Model Resolution Matrix). *Let us consider the following eigenvalue decomposition:*

$$\mathbf{Q}_{\text{dof}} \equiv \mathbf{B}^{1/2}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{H}\mathbf{B}^{1/2} = \mathbf{V}^T\mathbf{\Sigma}\mathbf{V}, \quad (35)$$

where \mathbf{V} is the matrix whose columns are the eigenvectors of \mathbf{Q}_{dof} $\{\mathbf{v}_i, i = 1, \dots, n\}$, and $\mathbf{\Sigma}$ is a diagonal matrix whose elements are the eigenvalues of \mathbf{Q}_{dof} $\{\sigma_i, i = 1, \dots, n\}$. The vectors $\{\mathbf{B}^{1/2}\mathbf{v}_i, i = 1, \dots, n\}$ form a basis of eigenvectors for the model resolution matrix \mathbf{A} .

Proof. Using (9), one can write $\mathbf{A} = \mathbf{B}^{1/2} \left(\mathbf{B}^{1/2}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{H}\mathbf{B}^{1/2} \right) \mathbf{B}^{-1/2}$. Therefore, $\mathbf{A} = \mathbf{B}^{1/2}\mathbf{V}^T\mathbf{\Sigma}\mathbf{V}\mathbf{B}^{-1/2} = \mathbf{B}^{1/2}\mathbf{V}^T\mathbf{\Sigma}(\mathbf{B}^{1/2}\mathbf{V}^T)^{-1}$, and the vectors $\{\mathbf{B}^{1/2}\mathbf{v}_i, i = 1, \dots, n\}$ diagonalize \mathbf{A} . \square

Theorem 2.13 (Maximum-DOFS Projection). *Let us define \mathbf{V}_k the matrix whose columns are the first k eigenvectors of \mathbf{Q}_{dof} $\{\mathbf{v}_i, i = 1, \dots, k\}$, and let us define the projector:*

$$\mathbf{\Pi}_{\text{dof}} = \mathbf{B}^{1/2}\mathbf{V}_k\mathbf{V}_k^T\mathbf{B}^{-1/2} \quad (36)$$

The projector $\mathbf{\Pi}_{\text{dof}}$ maximizes the DOFS of the projected Bayesian problem $\mathcal{B}_\Pi = (E, F, \mathbf{H}\mathbf{\Pi}, \mathbf{B}, \mathbf{R}_\Pi)$ among all projectors $\mathbf{\Pi}$ of maximum rank k , i.e.:

$$\forall \mathbf{\Pi} \in \mathcal{P}, \text{Tr}(\mathbf{A}_{\mathbf{\Pi}_{\text{dof}}}) \geq \text{Tr}(\mathbf{A}_\Pi), \quad (37)$$

where \mathbf{A}_Π is the model resolution matrix associated with the problem \mathcal{B}_Π .

Proof. Let us consider a projection associated with an optimal prolongation, i.e., of the form $\Pi = \mathbf{B}^{1/2} \mathbf{U} \mathbf{U}^T \mathbf{B}^{-1/2}$, with \mathbf{U} an orthogonal matrix (see (34)). Replacing Π by this expression in (19) yields $\mathbf{A}_\Pi = \mathbf{B}^{1/2} \mathbf{U} \mathbf{U}^T \mathbf{B}^{1/2} \mathbf{H}^T (\mathbf{H} \mathbf{B}^T \mathbf{H}^T + \mathbf{R})^{-1} \mathbf{H} \mathbf{B}^{1/2} \mathbf{U} \mathbf{U}^T \mathbf{B}^{-1/2}$. Using the property of invariance of the trace under matrix permutation and the fact that $\mathbf{U}^T \mathbf{U} = \mathbf{Id}_k$, one obtains $\text{Tr}(\mathbf{A}_\Pi) = \text{Tr}(\mathbf{U}^T \mathbf{B}^{1/2} \mathbf{H}^T (\mathbf{H} \mathbf{B}^T \mathbf{H}^T + \mathbf{R})^{-1} \mathbf{H} \mathbf{B}^{1/2} \mathbf{U})$. By Lemma .2, the maximum of $\text{Tr}(\mathbf{A}_\Pi)$ is obtained for $\mathbf{U} = \mathbf{V}$, where \mathbf{V} is the matrix whose columns are the first k eigenvectors of \mathbf{Q}_{dof} , $\{\mathbf{v}_i, i = 1, \dots, k\}$, which proves (37). \square

Remarks 2.6. As suggested in 2.1.2, another criteria to optimize the projection is to minimize the total error variance (i.e., $\text{Tr}(\mathbf{P}_\Pi^a)$). However, unlike the maximum-DOFS projection, this minimum-error projection does not have a simple analytical expression, which prevents its efficient computation. In Section 2.3, we present alternative optimal approximations of the posterior error covariance matrix whose Frobenius distance to the true posterior error covariance matrix is minimal and whose total error variance is closest to true total error variance.

Remarks 2.7. The optimal projection defined in Thm. 2.13 has been proposed by Spantini *et al.* (2015). We note that in their study the projected problem is defined as $\mathcal{B}_{\Pi_{\text{opt}}} = (E, F, \mathbf{H} \Pi_{\text{opt}}, \mathbf{B}, \mathbf{R})$. However, as discussed in the present study, it is necessary to include a representativeness error, i.e., to use $\mathbf{R}_{\Pi_{\text{opt}}}$ instead of \mathbf{R} when defining the projected Bayesian problem. Spantini *et al.* (2015) overlooked this issue in their analysis, which is taken into account in our proofs.

Once the truncated eigendecomposition of \mathbf{Q}_{dof} is available, the posterior mean and posterior error covariance of the projected problem can be explicitly expressed as a function of the first k eigenvectors and eigenvalues. Note that in its current form \mathbf{Q}_{dof} requires the inversion of a potentially high-dimensional $p \times p$ matrix. In fact, one can circumvent this difficulty by noting that the eigendecomposition of \mathbf{Q}_{dof} can be efficiently obtained from the eigendecomposition of an auxiliary matrix called the *prior-preconditioned Hessian*. The following properties establish the formulas to compute the maximum-DOFS solution based on that improved implementation:

Proposition 2.14 (Posterior Solution of the maximum-DOFS Projection). *Let us define the prior-preconditioned Hessian $\widehat{\mathbf{H}}_p \equiv \mathbf{B}^{1/2} \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \mathbf{B}^{1/2}$ and its eigenvalue decomposition $\widehat{\mathbf{H}}_p = \mathbf{V}'^T \mathbf{\Lambda} \mathbf{V}'$. One has:*

$$\mathbf{V} = \mathbf{V}' \quad (38)$$

$$\mathbf{\Sigma} = \mathbf{\Lambda} (\mathbf{Id} + \mathbf{\Lambda})^{-1}, \quad (39)$$

where $\mathbf{Q}_{\text{dof}} = \mathbf{V}^T \mathbf{\Sigma} \mathbf{V}$ is the eigendecomposition of \mathbf{Q}_{dof} . Moreover, the solution of the maximum-DOFS projection can be expressed as:

$$\mathbf{x}_{\Pi_{\text{dof}}}^a = \mathbf{B}^{1/2} \left(\sum_{i=1}^k \lambda_i^{1/2} (1 + \lambda_i)^{-1} \mathbf{v}_i \mathbf{w}_i^T \right) \mathbf{R}^{-1/2} \mathbf{d} \quad (40)$$

$$\mathbf{P}_{\Pi_{\text{dof}}}^a = \mathbf{B}^{1/2} \sum_{i=1}^k (1 + \lambda_i)^{-1} \mathbf{v}_i \mathbf{v}_i^T \mathbf{B}^{1/2} \quad (41)$$

$$\mathbf{A}_{\Pi_{\text{dof}}} = \mathbf{B}^{1/2} \left(\sum_{i=1}^k \lambda_i (1 + \lambda_i)^{-1} \mathbf{v}_i \mathbf{v}_i^T \right) \mathbf{B}^{-1/2}, \quad (42)$$

where $\mathbf{w}_i = \mathbf{R}^{-1/2} \mathbf{H} \mathbf{B}^{1/2} \mathbf{v}_i$ and the $\{\lambda_i, i = 1, \dots, n\}$ are the diagonal elements of $\mathbf{\Lambda}$.

Proof. Let us first prove (38)-(39). The matrix \mathbf{Q}_{dof} can be rewritten:

$$\mathbf{Q}_{\text{dof}} = \mathbf{B}^{1/2} \mathbf{H}^T \mathbf{R}^{-1/2} (\mathbf{R}^{-1/2} \mathbf{H} \mathbf{B}^T \mathbf{H}^T \mathbf{R}^{-1/2} + \mathbf{Id})^{-1} \mathbf{R}^{-1/2} \mathbf{H} \mathbf{B}^{1/2} \quad (43)$$

$$= \mathbf{B}^{1/2} \mathbf{H}^T \mathbf{R}^{-1/2} \mathbf{W}^T (\mathbf{Id} - \mathbf{\Lambda} (\mathbf{Id} + \mathbf{\Lambda})^{-1}) \mathbf{W} \mathbf{R}^{-1/2} \mathbf{H} \mathbf{B}^{1/2}, \quad (44)$$

where $\mathbf{B}^{1/2}\mathbf{H}^T\mathbf{R}^{-1/2} = \mathbf{V}^T\mathbf{\Lambda}^{1/2}\mathbf{W}$ is the SVD of the square-root of the prior-preconditioned matrix $\widehat{\mathbf{H}}_p$ and the Shermann-Morrison-Woodbury formula was applied to derive $(\mathbf{R}^{-1/2}\mathbf{H}\mathbf{B}\mathbf{H}^T\mathbf{R}^{-1/2} + \mathbf{Id})^{-1} = \mathbf{W}^T(\mathbf{Id} - \mathbf{\Lambda}(\mathbf{Id} + \mathbf{\Lambda})^{-1})\mathbf{W}$. Replacing the square-root of $\mathbf{B}^{1/2}\mathbf{H}^T\mathbf{R}^{-1/2}$ by its SVD in (44) and using the fact that $\mathbf{W}\mathbf{W}^T = \mathbf{Id}$, one obtains:

$$\mathbf{Q}_{\text{dof}} = \mathbf{V}^T \mathbf{\Lambda} (\mathbf{I} + \mathbf{\Lambda})^{-1} \mathbf{V}$$

To prove (40)-(42), we first use formulas (3), (6), and (9) for \mathbf{x}^a , \mathbf{P}^a and \mathbf{A} , respectively, and substitute \mathbf{Q}_{dof} in them to obtain the following expressions:

$$\mathbf{x}^a = \mathbf{B}^{1/2}(\mathbf{Id} - \mathbf{Q}_{\text{dof}})\mathbf{B}^{1/2}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{d} \quad (45)$$

$$\mathbf{P}^a = \mathbf{B}^{1/2}(\mathbf{Id} - \mathbf{Q}_{\text{dof}})\mathbf{B}^{1/2} \quad (46)$$

$$\mathbf{A} = \mathbf{B}^{1/2}\mathbf{Q}_{\text{dof}}\mathbf{B}^{-1/2} \quad (47)$$

We then substitute those expressions in formulas (31)-(33) and replace $\mathbf{\Pi}$ by its optimal solution $\mathbf{\Pi}_{\text{dof}} = \mathbf{B}^{1/2}\mathbf{V}_k\mathbf{V}_k^T\mathbf{B}^{-1/2}$, which yields:

$$\mathbf{x}_{\mathbf{\Pi}_{\text{dof}}}^a = \mathbf{B}^{1/2}\mathbf{V}_k\mathbf{V}_k^T(\mathbf{Id} - \mathbf{Q}_{\text{dof}})\mathbf{B}^{1/2}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{d} \quad (48)$$

$$\mathbf{P}_{\mathbf{\Pi}_{\text{dof}}}^a = \mathbf{B}^{1/2}\mathbf{V}_k\mathbf{V}_k^T(\mathbf{Id} - \mathbf{Q}_{\text{dof}})\mathbf{V}_k\mathbf{V}_k^T\mathbf{B}^{1/2} \quad (49)$$

$$\mathbf{A}_{\mathbf{\Pi}_{\text{dof}}} = \mathbf{B}^{1/2}\mathbf{V}_k\mathbf{V}_k^T\mathbf{Q}_{\text{dof}}\mathbf{V}_k\mathbf{V}_k^T\mathbf{B}^{-1/2} \quad (50)$$

Noting $\mathbf{B}^{1/2}\mathbf{H}^T\mathbf{R}^{-1} = \mathbf{V}^T\mathbf{\Lambda}^{1/2}\mathbf{W}\mathbf{R}^{-1/2}$ in (48), and replacing \mathbf{Q}_{dof} by its SVD in (48)-(50), one obtains the desired formulas (40)-(42). \square

Note that an alternative formula can be derived for Eq. (40), which has the advantage that it does not require computation of the singular vectors $\{\mathbf{w}_i\}$:

Proposition 2.15 (Alternative Formulation for Posterior Mean of Maximum-DOFS Projection). *Using the previous notations, let $\{(\mathbf{v}_i, \lambda_i), i = 1, \dots, k\}$ be the first k eigenpairs of the prior-preconditioned Hessian $\widehat{\mathbf{H}}_p$. The posterior mean of the rank- k maximum-DOFS projection can be expressed as:*

$$\mathbf{x}_{\mathbf{\Pi}_{\text{dof}}}^a = \mathbf{B}^{1/2} \left[\sum_{i=1}^k (1 + \lambda_i)^{-1} \mathbf{v}_i \mathbf{v}_i^T \right] \mathbf{B}^{1/2} \mathbf{H}^T \mathbf{R}^{-1} \mathbf{d} \quad (51)$$

Proof. One has:

$$\mathbf{B}^{1/2} \left[\sum_{i=1}^k (1 + \lambda_i)^{-1} \mathbf{v}_i \mathbf{v}_i^T \right] \mathbf{B}^{1/2} \mathbf{H}^T \mathbf{R}^{-1} \mathbf{d} = \mathbf{B}^{1/2} \left[\sum_{i=1}^k (1 + \lambda_i)^{-1} \mathbf{v}_i \mathbf{v}_i^T \right] \mathbf{B}^{1/2} \mathbf{H}^T \mathbf{R}^{-1/2} \mathbf{R}^{-1/2} \mathbf{d} \quad (52)$$

$$= \mathbf{B}^{1/2} \left[\sum_{i=1}^k (1 + \lambda_i)^{-1} \mathbf{v}_i \mathbf{v}_i^T \right] \mathbf{V} \mathbf{\Lambda}^{1/2} \mathbf{W}^T \mathbf{R}^{-1/2} \mathbf{d} \quad (53)$$

$$= \mathbf{B}^{1/2} \left(\sum_{i=1}^k \lambda_i^{1/2} (1 + \lambda_i)^{-1} \mathbf{v}_i \mathbf{w}_i^T \right) \mathbf{R}^{-1/2} \mathbf{d}, \quad (54)$$

with:

$$\mathbf{w}_i = \mathbf{R}^{-1/2} \mathbf{H} \mathbf{B}^{1/2} \mathbf{v}_i$$

From (40), we obtain equality (51). \square

Finally, from Prop. 2.14, one also obtained the following useful result:

Corollary 2.16. *The DOFS of the rank- k projected Bayesian problem $\mathcal{B}_{\Pi_{opt}}$ with maximum DOFS is the sum of the first k eigenvalues of the model resolution matrix, i.e.:*

$$\text{Tr}(\mathbf{A}_{\Pi_{dof}}) = \sum_{i=1}^k \lambda_i (1 + \lambda_i)^{-1} \quad (55)$$

Proof.

$$\begin{aligned} \text{Tr}(\mathbf{A}_{\Pi_{dof}}) &= \text{Tr}(\mathbf{B}^{1/2} \left(\sum_{i=1}^k \lambda_i (1 + \lambda_i)^{-1} \mathbf{v}_i \mathbf{v}_i^T \right) \mathbf{B}^{-1/2}) \\ &= \text{Tr}(\mathbf{B}^{-1/2} \mathbf{B}^{1/2} \left(\sum_{i=1}^k \lambda_i (1 + \lambda_i)^{-1} \mathbf{v}_i \mathbf{v}_i^T \right)) \\ &= \sum_{i=1}^k \lambda_i (1 + \lambda_i)^{-1} \end{aligned}$$

□

2.2.5 Interpretation

Information Content of Subspaces Using our previous analysis, a natural generalization of the concept of information content to subspaces of a linear Bayesian problem can be derived. Given a subspace of dimension k defined by the basis $\{\mathbf{r}_i, i = 1, \dots, k\}$ and its associated matrix column \mathbf{R} , let us define the projection Π_R with range \mathbf{R} and direction $\mathbf{D} = \mathbf{Id} - \mathbf{B}^{-1}\mathbf{R}$, that is, $\Pi_R = \mathbf{R}(\mathbf{R}^T \mathbf{B}^{-1} \mathbf{R})^{-1} \mathbf{R}^T \mathbf{B}^{-1}$. One can verify that Π_R belongs to the class of optimal projections \mathcal{P}_{opt} defined in Prop. 2.11 by defining $\mathbf{U} = \mathbf{B}^{-1/2} \mathbf{R}(\mathbf{R}^T \mathbf{B}^{-1} \mathbf{R})^{-1/2}$, and noting that $\mathbf{U}^T \mathbf{U} = \mathbf{Id}$ and $\Pi_R = \mathbf{B}^{1/2} \mathbf{U} \mathbf{U}^T \mathbf{B}^{-1/2}$. With this particular choice for the direction of the projection Π_R , the information content of the subspace $\{\mathbf{r}_i, i = 1, \dots, k\}$ can be defined as the DOFS of the projected Bayesian problem \mathcal{B}_{Π_R} , that is, the DOFS of the Bayesian problem projected onto that subspace along the direction that maximizes the DOFS (see Thm 2.6).

Most Informed Subspaces Thm. 2.13 shows that the maximum-DOFS projection is constructed incrementally using $\Pi^k = \Pi^{k-1} + \mathbf{B}^{1/2} \mathbf{v}_i \mathbf{v}_i^T \mathbf{B}^{-1/2}$. Therefore, the subspace defined by the basis $\{\mathbf{B}^{1/2} \mathbf{v}_i, i = 1, \dots, k\}$, which corresponds to the range of Π^k , can be interpreted as the *most informed subspace* of dimension k , while the vector $\mathbf{B}^{1/2} \mathbf{v}_j$ defines the j th most constrained direction.

Independently Constrained Modes The vectors $\mathbf{B}^{1/2} \mathbf{v}_j$ are the eigenvectors of the model resolution matrix $\mathbf{A} \equiv \frac{\partial \mathbf{x}^a}{\partial \mathbf{x}^t}$. They can therefore be interpreted as the modes that are independently constrained by the observations, since one has (Eq. (42)) $\frac{\partial (\mathbf{B}^{1/2} \mathbf{v}_i)^a}{\partial (\mathbf{B}^{1/2} \mathbf{v}_j)^t} = \lambda_i (1 + \lambda_i)^{-1} \delta_{ij}$ (where δ_{ij} represents the Kronecker delta).

Projected Forward Model Based on Rem. 2.4, one can establish a link between the posterior solutions of the maximum-DOFS projection and the posterior solutions of the Bayesian problem $\mathcal{B}_{\Pi_{dof}} \equiv (E, F, \mathbf{H} \Pi_{dof}, \mathbf{B}, \mathbf{R}_{\Pi_{dof}})$, which corresponds to the initial Bayesian problem, \mathcal{B} , with a

projected forward model. Indeed, one can verify that the posterior solutions of $\mathcal{B}_{\mathbf{H}_{\Pi_{\text{dof}}}}$ can be expressed as:

$$\mathbf{x}_{\mathbf{H}_{\Pi_{\text{dof}}}}^a = \mathbf{x}_{\Pi_{\text{dof}}}^a \quad (56)$$

$$\mathbf{P}_{\mathbf{H}_{\Pi_{\text{dof}}}}^a = \mathbf{P}_{\Pi_{\text{dof}}}^a + \mathbf{B}^{1/2} \left(\sum_{i=k+1}^n \mathbf{v}_i \mathbf{v}_i^T \right) \mathbf{B}^{1/2} = \mathbf{B} - \mathbf{B}^{1/2} \left(\sum_{i=1}^k \lambda_i (1 + \lambda_i)^{-1} \mathbf{v}_i \mathbf{v}_i^T \right) \mathbf{B}^{1/2} \quad (57)$$

$$\mathbf{A}_{\mathbf{H}_{\Pi_{\text{dof}}}} = \mathbf{A}_{\Pi_{\text{dof}}} \quad (58)$$

The optimality properties of the low-rank update approximation (57) will be described and exploited in the following Section.

2.3 Link With Low-Rank Approximations

The maximum-DOFS projection constructed in Section 2.2.4 defines a rank- k Bayesian inverse problem whose information content is maximal among all rank- k projections of the initial Bayesian problem. Furthermore, the posterior mean and posterior error covariance matrix of the maximum-DOFS projection are also low-rank approximations to the initial full-dimensional posterior mean and posterior error covariance matrix, respectively. In this Section, we provide important optimality results associated with those low-rank approximations to the posterior solution. Additionally, useful optimality results associated with low-rank approximations to the posterior solution that do not correspond to projections are also provided, which can be alternatively used when only best approximations to the solution of the original Bayesian problem are sought and consistency between the approximated posterior mean and posterior errors is not required (see Section 2.1.3).

2.3.1 Optimal Low-Rank Approximations

Before establishing optimality results associated with low-rank approximations to the posterior solution, one needs to define an appropriate class of approximations. As discussed in Section 2.1.2, a natural class of approximations for the posterior error covariance matrix is the one that corresponds to negative updates to the prior error covariance matrix. This particular class is central to our analysis, since the negative update can be interpreted as the information content of the inversion (see Section 2.1.2). We note that previous studies have already demonstrated the importance of this approximation class for metrics useful in the Bayesian framework (e.g., Spantini *et al.* (2015); Cui *et al.* (2014)).

Definition 2.17 (Classes of Approximations). Let us define the following classes of matrices:

$$\begin{aligned} \mathcal{A}_k &\equiv \{\mathbf{M} \in \mathcal{M}_n \mid \text{rank}(\mathbf{M}) \leq k\} \\ \hat{\mathcal{A}}_k &\equiv \{\mathbf{M} \in \mathcal{M}_n \mid \mathbf{M} = \mathbf{B} - \mathbf{Q}\mathbf{Q}^T \geq 0, \text{rank}(\mathbf{Q}) \leq k\} \\ \hat{\mathcal{O}}_k &\equiv \{\mathbf{M} = \mathbf{P}\mathbf{H}^T\mathbf{R}^{-1} \mid \mathbf{P} \in \hat{\mathcal{A}}_k\}, \end{aligned}$$

where $\hat{\mathcal{A}}_k$ defines the class of negative semidefinite updates to the prior error covariance matrix \mathbf{B} .

The approximations that belong to the class \mathcal{A}_k are referred to as low-rank approximations, while the approximations that belong to the class $\hat{\mathcal{A}}_k$ are low-rank update approximations. Therefore, the class $\hat{\mathcal{O}}_k$ is associated with full-rank approximations to the posterior mean update.

In the following Proposition, optimality results for several posterior error covariance matrix approximations are provided. All approximations are based on the classes defined in Def. 2.17 and on the truncated eigendecomposition of either \mathbf{Q}_{dof} (see Section 2.2.4) or $\mathbf{Q}_{\text{var}} \equiv \mathbf{B} - \mathbf{P}^a$, which are both related to the information content of the inversion.

Proposition 2.18 (Optimal Approximations of the Posterior Error Covariance).

Using the previous notations, let us define:

$$\mathbf{P}_{\mathbf{H}_{\Pi_{\text{dof}}}}^a = \mathbf{B} - \mathbf{B}^{1/2} \left(\sum_{i=1}^k \lambda_i (1 + \lambda_i)^{-1} \mathbf{v}_i \mathbf{v}_i^T \right) \mathbf{B}^{1/2} \quad (59)$$

$$\mathbf{P}_{\text{var}}^a \equiv \mathbf{B} - \sum_{i=1}^k \delta_i \mathbf{u}_i \mathbf{u}_i^T, \quad (60)$$

where \mathbf{u}_i and δ_i are the i th eigenvector and eigenvalue, respectively, of $\mathbf{Q}_{\text{var}} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{H}\mathbf{B}$, and where $\mathbf{P}_{\mathbf{H}_{\Pi_{\text{dof}}}}^a$ is the posterior error covariance of the projected Bayesian problem $\mathcal{B}_{\mathbf{H}_{\Pi_{\text{opt}}}}$.

One has the following optimality properties:

$$\|\mathbf{P}_{\text{var}}^a - \mathbf{P}^a\|_F = \min_{\tilde{\mathbf{P}} \in \hat{\mathcal{A}}_k} \|\tilde{\mathbf{P}} - \mathbf{P}^a\|_F = \sqrt{\sum_{i>k} \delta_i^2} \quad (61)$$

$$\|\mathbf{P}_{\mathbf{H}_{\Pi_{\text{dof}}}}^a - \mathbf{P}^a\|_{F, \mathbf{B}^{-1}} = \min_{\tilde{\mathbf{P}} \in \hat{\mathcal{A}}_k} \|\tilde{\mathbf{P}} - \mathbf{P}^a\|_{F, \mathbf{B}^{-1}} = \sqrt{\sum_{i>k} \left(\frac{\lambda_i}{1 + \lambda_i} \right)^2} \quad (62)$$

$$\|\mathbf{P}_{\mathbf{H}_{\Pi_{\text{dof}}}}^a - \mathbf{P}^a\|_{F, (\mathbf{P}^a)^{-1}} = \min_{\tilde{\mathbf{P}} \in \hat{\mathcal{A}}_k} \|\tilde{\mathbf{P}} - \mathbf{P}^a\|_{F, (\mathbf{P}^a)^{-1}} = \sqrt{\sum_{i>k} \lambda_i^2} \quad (63)$$

$$\|\mathbf{P}_{\Pi_{\text{dof}}}^a - \mathbf{P}^a\|_{F, \mathbf{B}^{-1}} = \min_{\tilde{\mathbf{P}} \in \mathcal{A}_k} \|\tilde{\mathbf{P}} - \mathbf{P}^a\|_{F, \mathbf{B}^{-1}} = \sqrt{\sum_{i>k} \left(\frac{1}{1 + \lambda_i} \right)^2} \quad (64)$$

$$\|\mathbf{P}_{\Pi_{\text{dof}}}^a - \mathbf{P}^a\|_{F, (\mathbf{P}^a)^{-1}} = \min_{\tilde{\mathbf{P}} \in \mathcal{A}_k} \|\tilde{\mathbf{P}} - \mathbf{P}^a\|_{F, (\mathbf{P}^a)^{-1}} = \sqrt{n - k} \quad (65)$$

where:

- $\|\cdot\|_F$ represents the Frobenius norm.
- $\|\cdot\|_{F, \mathbf{W}}$ is the weighted Frobenius norm defined by $\|\mathbf{M}\|_{F, \mathbf{W}} = \|\mathbf{B}^{1/2} \mathbf{M} \mathbf{B}^{1/2}\|_F$, where $\mathbf{B}^{1/2}$ is a square-root of \mathbf{W} ($\mathbf{W} = \mathbf{L}\mathbf{L}^T$).

Proof. The proof of (61) follows immediately from the Eckart-Young theorem [Eckart and Young, 1936], since one has $\|\mathbf{P}_{\text{var}}^a - \mathbf{P}^a\|_F = \|\sum_{i=1}^k \delta_i \mathbf{u}_i \mathbf{u}_i^T - \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{H}\mathbf{B}\|_F = \min_{\tilde{\mathbf{P}} \in \hat{\mathcal{A}}_k} \|\tilde{\mathbf{P}} - \mathbf{Q}_{\text{var}}\|_F$.

The proofs for formulas (62)-(65) are obtained by writing:

$$\mathbf{P}_{\mathbf{H}_{\Pi_{\text{dof}}}}^a - \mathbf{P}^a = \mathbf{B}^{1/2} \left(\sum_{i=1}^k \lambda_i (1 + \lambda_i)^{-1} \mathbf{v}_i \mathbf{v}_i^T - \sum_{i=1}^n \lambda_i (1 + \lambda_i)^{-1} \mathbf{v}_i \mathbf{v}_i^T \right) \mathbf{B}^{1/2} \quad (66)$$

$$\mathbf{P}_{\Pi_{\text{dof}}}^a - \mathbf{P}^a = \mathbf{B}^{1/2} \left(\sum_{i=1}^k (1 + \lambda_i)^{-1} \mathbf{v}_i \mathbf{v}_i^T - \sum_{i=1}^n (1 + \lambda_i)^{-1} \mathbf{v}_i \mathbf{v}_i^T \right) \mathbf{B}^{1/2} \quad (67)$$

Multiplying on the left and on the right by the inverse of the square-root of \mathbf{B} in (66) and (67), we obtain:

$$\begin{aligned}
\|\mathbf{P}_{\mathbf{H}\Pi_{\text{dof}}}^a - \mathbf{P}^a\|_{F, \mathbf{B}^{-1}} &= \left\| \left(\sum_{i=1}^k \lambda_i (1 + \lambda_i)^{-1} \mathbf{v}_i \mathbf{v}_i^T - \sum_{i=1}^n \lambda_i (1 + \lambda_i)^{-1} \mathbf{v}_i \mathbf{v}_i^T \right) \right\|_F \\
&= \min_{\tilde{\mathbf{M}} \in \tilde{\mathcal{A}}_k} \left\| \tilde{\mathbf{M}} - \sum_{i=1}^n \lambda_i (1 + \lambda_i)^{-1} \mathbf{v}_i \mathbf{v}_i^T \right\|_F \\
&= \sqrt{\sum_{i>k} \left(\frac{\lambda_i}{1 + \lambda_i} \right)^2} \\
\|\mathbf{P}_{\Pi_{\text{dof}}}^a - \mathbf{P}^a\|_{F, \mathbf{B}^{-1}} &= \left\| \left(\sum_{i=1}^k (1 + \lambda_i)^{-1} \mathbf{v}_i \mathbf{v}_i^T - \sum_{i=1}^n (1 + \lambda_i)^{-1} \mathbf{v}_i \mathbf{v}_i^T \right) \right\|_F \\
&= \min_{\tilde{\mathbf{M}} \in \tilde{\mathcal{A}}_k} \left\| \tilde{\mathbf{M}} - \sum_{i=1}^n (1 + \lambda_i)^{-1} \mathbf{v}_i \mathbf{v}_i^T \right\|_F \\
&= \sqrt{\sum_{i>k} \left(\frac{1}{1 + \lambda_i} \right)^2}
\end{aligned}$$

Using the previous notations, a square-root of $(\mathbf{P}^a)^{-1} = (\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}) = \mathbf{B}^{-1/2} (\mathbf{Id} + \mathbf{B}^{1/2} \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \mathbf{B}^{1/2}) \mathbf{B}^{-1/2}$ is given by $\mathbf{L}_{\mathbf{P}^a-1} = \mathbf{B}^{-1/2} \mathbf{V} (\mathbf{Id} + \mathbf{\Lambda})^{1/2} \mathbf{V}^T$. Multiplying on the left and on the right by $\mathbf{L}_{\mathbf{P}^a}^T$ and $\mathbf{L}_{\mathbf{P}^a}$, respectively, in (66) and (67), we obtain:

$$\begin{aligned}
\|\mathbf{P}_{\mathbf{H}\Pi_{\text{dof}}}^a - \mathbf{P}^a\|_{F, (\mathbf{P}^a)^{-1}} &= \left\| \left(\sum_{i=1}^k \lambda_i \mathbf{v}_i \mathbf{v}_i^T - \sum_{i=1}^n \lambda_i \mathbf{v}_i \mathbf{v}_i^T \right) \right\|_F \\
&= \min_{\tilde{\mathbf{M}} \in \tilde{\mathcal{A}}_k} \left\| \tilde{\mathbf{M}} - \sum_{i=1}^n \lambda_i \mathbf{v}_i \mathbf{v}_i^T \right\|_F \\
&= \sqrt{\sum_{i>k} \lambda_i^2} \\
\|\mathbf{P}_{\Pi_{\text{dof}}}^a - \mathbf{P}^a\|_{F, (\mathbf{P}^a)^{-1}} &= \left\| \left(\sum_{i=1}^k \mathbf{v}_i \mathbf{v}_i^T - \sum_{i=1}^n \mathbf{v}_i \mathbf{v}_i^T \right) \right\|_F \\
&= \min_{\tilde{\mathbf{M}} \in \tilde{\mathcal{A}}_k} \left\| \tilde{\mathbf{M}} - \sum_{i=1}^n \mathbf{v}_i \mathbf{v}_i^T \right\|_F \\
&= \sqrt{n - k}
\end{aligned}$$

□

Remarks 2.8. It has been shown in Spantini *et al.* (2015) that $\mathbf{P}_{\mathbf{H}\Pi_{\text{dof}}}^a$ also verifies: $d_{\mathcal{F}}(\mathbf{P}_{\mathbf{H}\Pi_{\text{dof}}}^a, \mathbf{P}^a) = \min_{\tilde{\mathbf{P}} \in \tilde{\mathcal{A}}_k} d_{\mathcal{F}}(\tilde{\mathbf{P}}, \mathbf{P}^a)$, where $d_{\mathcal{F}}$ is the *Förstner distance*, defined by $d_{\mathcal{F}}(\mathbf{P}, \mathbf{N}) = \sum_i (\ln \sigma_i)^2$, where (σ_i) is the sequence of generalized eigenvalues of the pencil (\mathbf{P}, \mathbf{N}) .

We now establish optimality results for several posterior mean approximations. In addition to the maximum-DOFS solution $(\mathbf{x}_{\Pi_{\text{dof}}}^a)$, a full-rank posterior mean approximation $(\mathbf{x}_{\text{FRdof}}^a)$ is considered. It is obtained by replacing the posterior error covariance implicit in Eq. (3) by the low-rank update approximation $\mathbf{P}_{\mathbf{H}\Pi_{\text{dof}}}^a$. The truncated eigendecomposition of \mathbf{Q}_{var} is also

exploited to define another optimal approximation of the posterior mean ($\mathbf{x}_{\text{var}}^a$). The optimality results (72) and (74) below can be found in Spantini *et al.* (2015). We recall the proofs here since the same principles can be applied to prove the other optimality results.

Proposition 2.19 (Optimal Approximations of the Posterior Mean).

Using the previous notations, let us define the following posterior mean approximations:

$$\mathbf{x}_{\text{FRdof}}^a \equiv \mathbf{P}_{\mathbf{H}\Pi_{\text{dof}}}^a \mathbf{H}^T \mathbf{R}^{-1} \mathbf{d} \quad (68)$$

$$\mathbf{x}_{\text{var}}^a = \sum_{i=1}^k \delta_i^{1/2} \mathbf{u}_i \mathbf{z}_i^T \left(\mathbf{H} \mathbf{B} \mathbf{H}^T + \mathbf{R} \right)^{-1/2} \mathbf{d}, \quad (69)$$

where $\mathbf{z}_i = (\mathbf{H} \mathbf{B} \mathbf{H}^T + \mathbf{R})^{-1/2} \mathbf{H} \mathbf{B} \mathbf{u}_i$.

One has the following optimality properties:

$$\mathbb{E} \|\mathbf{x}_{\text{var}}^a - \mathbf{x}^a\|^2 = \min_{\tilde{\mathbf{K}} \in \mathcal{A}_k} \mathbb{E} \|(\tilde{\mathbf{K}} - \mathbf{K}) \mathbf{d}\|^2 = \sum_{i>k} \delta_i \quad (70)$$

$$\mathbb{E} \|\mathbf{x}_{\Pi_{\text{dof}}}^a - \mathbf{x}^a\|_{\mathbf{B}^{-1}}^2 = \min_{\tilde{\mathbf{K}} \in \mathcal{A}_k} \mathbb{E} \|(\tilde{\mathbf{K}} - \mathbf{K}) \mathbf{d}\|_{\mathbf{B}^{-1}}^2 = \sum_{i>k} \frac{\lambda_i}{1 + \lambda_i} \quad (71)$$

$$\mathbb{E} \|\mathbf{x}_{\Pi_{\text{dof}}}^a - \mathbf{x}^a\|_{(\mathbf{P}^a)^{-1}}^2 = \min_{\tilde{\mathbf{K}} \in \mathcal{A}_k} \mathbb{E} \|(\tilde{\mathbf{K}} - \mathbf{K}) \mathbf{d}\|_{(\mathbf{P}^a)^{-1}}^2 = \sum_{i>k} \lambda_i \quad (72)$$

$$\mathbb{E} \|\mathbf{x}_{\text{FRdof}}^a - \mathbf{x}^a\|_{\mathbf{B}^{-1}}^2 = \min_{\tilde{\mathbf{P}} \in \mathcal{O}_k} \mathbb{E} \|(\tilde{\mathbf{P}} - \mathbf{P}^a) \mathbf{H}^T \mathbf{R}^{-1} \mathbf{d}\|_{\mathbf{B}^{-1}}^2 = \sum_{i>k} \frac{\lambda_i^3}{1 + \lambda_i} \quad (73)$$

$$\mathbb{E} \|\mathbf{x}_{\text{FRdof}}^a - \mathbf{x}^a\|_{(\mathbf{P}^a)^{-1}}^2 = \min_{\tilde{\mathbf{P}} \in \mathcal{O}_k} \mathbb{E} \|(\tilde{\mathbf{P}} - \mathbf{P}^a) \mathbf{H}^T \mathbf{R}^{-1} \mathbf{d}\|_{(\mathbf{P}^a)^{-1}}^2 = \sum_{i>k} \lambda_i^3, \quad (74)$$

where:

- $\|\cdot\|$ represent the Euclidian norm.
- $\|\mathbf{x}\|_{\mathbf{W}} = \sqrt{\mathbf{x}^T \mathbf{W} \mathbf{x}}$ is the weighted Euclidian norm.
- \mathbb{E} is the average operator.
- $\mathbf{K} = \mathbf{B} \mathbf{H}^T (\mathbf{H} \mathbf{B} \mathbf{H}^T + \mathbf{R})^{-1}$ is the gain matrix of the initial Bayesian problem.

Proof. To prove (70), we use Lemma .3 and the fact that $\mathbb{E}(\mathbf{d} \mathbf{d}^T) = \mathbf{H} \mathbf{B} \mathbf{H}^T + \mathbf{R}$:

$$\mathbb{E} \|(\tilde{\mathbf{K}} - \mathbf{K}) \mathbf{d}\|^2 = \|(\tilde{\mathbf{K}} - \mathbf{K}) (\mathbf{H} \mathbf{B} \mathbf{H}^T + \mathbf{R})^{1/2}\|_F^2 \quad (75)$$

$$= \|(\tilde{\mathbf{K}} (\mathbf{H} \mathbf{B} \mathbf{H}^T + \mathbf{R})^{1/2} - \mathbf{B} \mathbf{H}^T (\mathbf{H} \mathbf{B} \mathbf{H}^T + \mathbf{R})^{-1/2})\|_F^2 \quad (76)$$

Now, using Theorem 2.1 of Friedland and Torokhti (2007), a solution of $\min_{\tilde{\mathbf{K}} \in \mathcal{A}_k} \mathbb{E} \|(\tilde{\mathbf{K}} - \mathbf{K}) \mathbf{d}\|^2$ is given by:

$$\tilde{\mathbf{K}}_{\text{opt}} = \sum_{i=1}^k \delta_i^{1/2} \mathbf{u}_i \mathbf{z}_i^T \left(\mathbf{H} \mathbf{B} \mathbf{H}^T + \mathbf{R} \right)^{-1/2}, \quad (77)$$

where $\sum_{i=1}^k \delta_i^{1/2} \mathbf{u}_i \mathbf{z}_i^T$ is the truncated SVD of rank k of $\mathbf{B} \mathbf{H}^T (\mathbf{H} \mathbf{B} \mathbf{H}^T + \mathbf{R})^{-1/2}$ and $(\mathbf{H} \mathbf{B} \mathbf{H}^T + \mathbf{R})^{1/2}$ is a non-singular square-root of $\mathbf{H} \mathbf{B} \mathbf{H}^T + \mathbf{R}$. Replacing $\tilde{\mathbf{K}}$ by (79) and $\mathbf{B} \mathbf{H}^T (\mathbf{H} \mathbf{B} \mathbf{H}^T + \mathbf{R})^{-1/2}$ by

$\sum_{i=1}^n \delta_i^{1/2} \mathbf{u}_i \mathbf{z}_i^T$ in (76) yields (70).

Below we prove (72). The proof for (71) is obtained similarly. To prove (72), we also use Lemma .3 and the square-roots $\mathbf{L}_{(\mathbf{P}^a)^{-1}} = \mathbf{B}^{-1/2} \mathbf{V}(\mathbf{I} + \mathbf{\Lambda})^{1/2} \mathbf{V}^T$ and $\mathbf{L}_{\mathbf{Y}} = \mathbf{R}^{1/2} \mathbf{W}(\mathbf{I} + \mathbf{\Lambda})^{1/2} \mathbf{W}^T$ of $(\mathbf{P}^a)^{-1}$ and $\mathbf{Y} = \mathbf{H} \mathbf{B} \mathbf{H}^T + \mathbf{R}$, respectively. One has:

$$\mathbb{E} \|\tilde{\mathbf{K}} - \mathbf{K}\|_{(\mathbf{P}^a)^{-1}}^2 = \|\mathbf{L}_{(\mathbf{P}^a)^{-1}}^T (\tilde{\mathbf{K}} - \mathbf{K}) \mathbf{L}_{\mathbf{Y}}\|_F^2 \quad (78)$$

Using Theorem 2.1 of Friedland and Torokhti (2007), a solution of $\min_{\tilde{\mathbf{K}} \in \mathcal{A}_k} \mathbb{E} \|\tilde{\mathbf{K}} - \mathbf{K}\|_{(\mathbf{P}^a)^{-1}}^2$ is therefore given by:

$$\tilde{\mathbf{K}}_{\text{opt}} = \mathbf{L}_{(\mathbf{P}^a)^{-1}}^{-T} \left[\mathbf{L}_{(\mathbf{P}^a)^{-1}}^T \mathbf{K} \mathbf{L}_{\mathbf{Y}} \right]_k \mathbf{L}_{\mathbf{Y}}^{-1}, \quad (79)$$

where $[\mathbf{M}]_k$ is the rank- k truncated SVD of the matrix \mathbf{M} . Further developing (79), one obtains:

$$\begin{aligned} \tilde{\mathbf{K}}_{\text{opt}} &= \mathbf{B}^{1/2} \mathbf{V}(\mathbf{I} + \mathbf{\Lambda})^{-1/2} \mathbf{V}^T \\ &\quad \left[\mathbf{V}(\mathbf{I} + \mathbf{\Lambda})^{1/2} \mathbf{V}^T \mathbf{B}^{1/2} \mathbf{H}^T \mathbf{R}^{-1/2} (\mathbf{I} + \mathbf{R}^{-1/2} \mathbf{H} \mathbf{B} \mathbf{H}^T \mathbf{R}^{-1/2})^{-1} \mathbf{W}(\mathbf{I} + \mathbf{\Lambda})^{1/2} \mathbf{W}^T \right]_k \\ &\quad \mathbf{W}(\mathbf{I} + \mathbf{\Lambda})^{-1/2} \mathbf{W}^T \mathbf{R}^{-1/2} \end{aligned}$$

Using $\mathbf{B}^{1/2} \mathbf{H}^T \mathbf{R}^{-1/2} = \mathbf{V} \mathbf{\Lambda}^{1/2} \mathbf{W}^T$ in the expression above one obtains:

$$\begin{aligned} \tilde{\mathbf{K}}_{\text{opt}} &= \mathbf{B}^{1/2} \mathbf{V}(\mathbf{I} + \mathbf{\Lambda})^{-1/2} \mathbf{V}^T \\ &\quad \left[\mathbf{V}(\mathbf{I} + \mathbf{\Lambda})^{1/2} \mathbf{V}^T \mathbf{V} \mathbf{\Lambda}^{1/2} \mathbf{W}^T \mathbf{W}(\mathbf{I} - \mathbf{\Lambda}(\mathbf{I} + \mathbf{\Lambda})^{-1}) \mathbf{W}^T \mathbf{W}(\mathbf{I} + \mathbf{\Lambda})^{1/2} \mathbf{W}^T \right]_k \\ &\quad \mathbf{W}(\mathbf{I} + \mathbf{\Lambda})^{-1/2} \mathbf{W}^T \mathbf{R}^{-1/2} \\ &= \mathbf{B}^{1/2} \mathbf{V}(\mathbf{I} + \mathbf{\Lambda})^{-1/2} \mathbf{V}^T \\ &\quad \left[\sum_{i=1}^k \mathbf{v}_i \mathbf{w}_i^T (1 + \lambda_i)^{1/2} \lambda_i^{1/2} (1 - \lambda_i (1 + \lambda_i)^{-1}) (1 + \lambda_i)^{1/2} \right] \\ &\quad \mathbf{W}(\mathbf{I} + \mathbf{\Lambda})^{-1/2} \mathbf{W}^T \mathbf{R}^{-1/2} \\ &= \mathbf{B}^{1/2} \left(\sum_{i=1}^k \lambda_i^{1/2} (1 + \lambda_i)^{-1} \mathbf{v}_i \mathbf{w}_i^T \right) \mathbf{R}^{-1/2}, \end{aligned}$$

where we used the orthogonality properties of \mathbf{W} and \mathbf{V} . Finally, using

$\tilde{\mathbf{K}}_{\text{opt}} = \mathbf{B}^{1/2} \left(\sum_{i=1}^k \lambda_i^{1/2} (1 + \lambda_i)^{-1} \mathbf{v}_i \mathbf{w}_i^T \right) \mathbf{R}^{-1/2}$ and $\mathbf{K} = \mathbf{B}^{1/2} \left(\sum_{i=1}^n \lambda_i^{1/2} (1 + \lambda_i)^{-1} \mathbf{v}_i \mathbf{w}_i^T \right) \mathbf{R}^{-1/2}$ in the right-hand side of (78) leads to $\mathbb{E} \|\tilde{\mathbf{K}}_{\text{opt}} - \mathbf{K}\|_{(\mathbf{P}^a)^{-1}}^2 = \sum_{i>k} \lambda_i$, which proves (72).

Below we prove (74), (73) being obtained similarly. Using expression (3) for the posterior mean and Lemma .3, we obtain:

$$\mathbb{E} \|\mathbf{x}_{\text{FR}_{\text{dof}}}^a - \mathbf{x}^a\|_{(\mathbf{P}^a)^{-1}}^2 = \mathbb{E} \|(\tilde{\mathbf{P}} - \mathbf{P}^a) \mathbf{H}^T \mathbf{R}^{-1} \mathbf{d}\|_{(\mathbf{P}^a)^{-1}}^2 \quad (80)$$

$$= \|\mathbf{L}_{(\mathbf{P}^a)^{-1}}^T (\tilde{\mathbf{P}} - \mathbf{P}^a) \mathbf{H}^T \mathbf{R}^{-1} \mathbf{L}_{\mathbf{Y}}\|_F^2 \quad (81)$$

We first note that:

$$\min_{\tilde{\mathbf{P}} \in \tilde{\mathcal{O}}_k} \|\mathbf{L}_{(\mathbf{P}^a)^{-1}}^T (\tilde{\mathbf{P}} - \mathbf{P}^a) \mathbf{H}^T \mathbf{R}^{-1} \mathbf{L}_{\mathbf{Y}}\|_F^2 = \min_{\tilde{\mathbf{F}} \in \tilde{\mathcal{A}}_k} \|\mathbf{L}_{(\mathbf{P}^a)^{-1}}^T (\tilde{\mathbf{F}} - (\mathbf{P}^a - \mathbf{B})) \mathbf{H}^T \mathbf{R}^{-1} \mathbf{L}_{\mathbf{Y}}\|_F^2$$

Using Theorem 2.1 of Friedland and Torokhti (2007), a solution of $\min_{\tilde{\mathbf{F}} \in \mathcal{A}_k} \|\mathbf{L}_{(\mathbf{P}^a)^{-1}}^T (\tilde{\mathbf{F}} - (\mathbf{P}^a - \mathbf{B})) \mathbf{H}^T \mathbf{R}^{-1} \mathbf{L}_Y\|_F^2$ is given by:

$$\tilde{\mathbf{F}}_{\text{opt}} = \mathbf{L}_{(\mathbf{P}^a)^{-1}}^{-T} \left[\mathbf{L}_{(\mathbf{P}^a)^{-1}}^T (\mathbf{P}^a - \mathbf{B}) \mathbf{H}^T \mathbf{R}^{-1} \mathbf{L}_Y \right]_k (\mathbf{H}^T \mathbf{R}^{-1} \mathbf{L}_Y)^+,$$

where $^+$ denotes the Moore-Penrose pseudoinverse. One can verify that another minimizer of (81) is:

$$\tilde{\mathbf{F}}'_{\text{opt}} = \mathbf{L}_{(\mathbf{P}^a)^{-1}}^{-T} \left[\mathbf{L}_{(\mathbf{P}^a)^{-1}}^T (\mathbf{P}^a - \mathbf{B}) \mathbf{H}^T \mathbf{R}^{-1} \mathbf{L}_Y \right]_k (\mathbf{B}^{1/2} \mathbf{H}^T \mathbf{R}^{-1} \mathbf{L}_Y)^+ \mathbf{B}^{1/2} \quad (82)$$

Factorizing the expression above using the square-root $\mathbf{B}^{1/2} \mathbf{H}^T \mathbf{R}^{-1/2}$ and its SVD like before, we obtain:

$$\tilde{\mathbf{F}}'_{\text{opt}} = \mathbf{B}^{1/2} \mathbf{V} (\mathbf{I} + \mathbf{\Lambda})^{-1/2} \mathbf{V}^T \quad (83)$$

$$\left[\sum_{i=1}^k \mathbf{v}_i \mathbf{w}_i^T (1 + \lambda_i)^{1/2} (\lambda_i (1 + \lambda_i)^{-1}) \lambda_i^{1/2} (1 + \lambda_i)^{1/2} \right] \quad (84)$$

$$\mathbf{W} (\mathbf{I} + \mathbf{\Lambda})^{-1/2} \mathbf{\Lambda}^{-1/2} \mathbf{V}^T \mathbf{B}^{1/2} \quad (85)$$

$$= \mathbf{B}^{1/2} \left(\sum_{i=1}^k \lambda_i (1 + \lambda_i)^{-1} \mathbf{v}_i \mathbf{v}_i^T \right) \mathbf{B}^{1/2} \quad (86)$$

Therefore:

$$\tilde{\mathbf{P}}'_{\text{opt}} = \mathbf{B} - \tilde{\mathbf{F}}'_{\text{opt}} = \mathbf{B} - \mathbf{B}^{1/2} \left(\sum_{i=1}^k \lambda_i (1 + \lambda_i)^{-1} \mathbf{v}_i \mathbf{v}_i^T \right) \mathbf{B}^{1/2}, \quad (87)$$

which proves the first equality of (74). Finally, using $\tilde{\mathbf{P}}'_{\text{opt}} = \mathbf{B} - \mathbf{B}^{1/2} \left(\sum_{i=1}^k \lambda_i (1 + \lambda_i)^{-1} \mathbf{v}_i \mathbf{v}_i^T \right) \mathbf{B}^{1/2}$ and $\mathbf{P}^a = \mathbf{B} - \mathbf{B}^{1/2} \left(\sum_{i=1}^n \lambda_i (1 + \lambda_i)^{-1} \mathbf{v}_i \mathbf{v}_i^T \right) \mathbf{B}^{1/2}$ in (81) we obtain the equality with the right-hand side of (74). \square

Remarks 2.9. The approximations associated with the eigendecomposition of \mathbf{Q}_{var} , i.e., $\mathbf{x}_{\text{var}}^a$ and $\mathbf{P}_{\text{var}}^a$, both correspond to optimal total (non-normalized) error variance approximations. Indeed, $\mathbf{P}_{\text{var}}^a$ is also the negative low-rank update to the prior error that best approximates the total error variance, i.e., $|\text{Tr}(\mathbf{P}_{\text{var}}^a - \mathbf{P}^a)| = \min_{\mathbf{P} \in \hat{\mathcal{A}}_k} |\text{Tr}(\mathbf{P} - \mathbf{P}^a)|$.

2.3.2 Interpretation and Application

Interpretation of the Norms The norms considered for the posterior error covariance approximations in Prop. 2.18 are all based on the Frobenius norm, which is defined as $\|\mathbf{A}\|_F = \sqrt{\sum_i |a_{ij}|^2}$ (where a_{ij} is the element of \mathbf{A} associated with the i th row and j th column), or alternatively, as $\|\mathbf{A}\|_F = \sqrt{\sum_i \sigma_i^2}$ (where σ_i represents the i th singular value of \mathbf{A}). Therefore, this norm accounts for all elements of the matrix in the approximation, or equivalently in the context of covariances matrices, accounts for variances in all directions (this is not the case of, e.g., the spectral norm $\|\mathbf{A}\|_S = \max_i \sigma_i$). Several norms in Prop. 2.18 are weighted Frobenius norms. In the case of covariance matrix approximations such as Eq. (62) to (65), those norms can be interpreted as total approximation errors normalized by the variances of the principal modes associated with the weight matrices. Indeed, one has $\|\mathbf{A}\|_{F, \mathbf{Q}^{-1}} = \|\mathbf{V}^T \mathbf{D}^{-1/2} \mathbf{V} \mathbf{A} \mathbf{V}^T \mathbf{D}^{-1/2} \mathbf{V}\|_F$, where $\mathbf{Q} = \mathbf{V}^T \mathbf{D} \mathbf{V}$ is the eigendecomposition of the Hermitian matrix \mathbf{Q} . The matrix $\mathbf{D}^{-1/2} \mathbf{V}^T \mathbf{A} \mathbf{V} \mathbf{D}^{-1/2}$ can be interpreted as the covariance matrix \mathbf{A} expressed in the basis of the principal components of \mathbf{Q} (i.e.,

the eigenvectors \mathbf{V}), and whose variances are normalized by the variance of the principal modes (defined by the diagonal elements of \mathbf{D}). The left and right products by \mathbf{V}^T and \mathbf{V} , respectively, transform the resulting matrix back into the original canonical basis. Therefore, the \mathbf{B}^{-1} -weighted Frobenius norm in Prop. 2.18 measures the relative approximation error in the posterior error covariance matrix with respect to the prior errors, while the $(\mathbf{P}^a)^{-1}$ -weighted Frobenius norm measures the relative approximation error in the posterior error covariance matrix with respect to the posterior errors.

A similar analysis can be performed to interpret the statistical approximation error in the posterior mean in Prop. 2.19. Indeed, one has:

$$\begin{aligned}
\mathbb{E}\|\mathbf{x}_{\text{approx}}^a - \mathbf{x}^a\|_{\mathbf{Q}^{-1}}^2 &= \mathbb{E}[(\mathbf{x}_{\text{approx}}^a - \mathbf{x}^a)^T \mathbf{Q}^{-1} (\mathbf{x}_{\text{approx}}^a - \mathbf{x}^a)] \\
&= \mathbb{E}[(\mathbf{V}(\mathbf{x}_{\text{approx}}^a - \mathbf{x}^a))^T \mathbf{D}^{-1} (\mathbf{V}(\mathbf{x}_{\text{approx}}^a - \mathbf{x}^a))] \\
&= \mathbb{E}(\text{Tr}[(\mathbf{V}(\mathbf{x}_{\text{approx}}^a - \mathbf{x}^a))^T \mathbf{D}^{-1} (\mathbf{V}(\mathbf{x}_{\text{approx}}^a - \mathbf{x}^a))]) \\
&= \mathbb{E}\left(\text{Tr}\left[(\mathbf{D}^{-1/2} \mathbf{V}(\mathbf{x}_{\text{approx}}^a - \mathbf{x}^a))(\mathbf{D}^{-1/2} \mathbf{V}(\mathbf{x}_{\text{approx}}^a - \mathbf{x}^a))^T\right]\right) \\
&= \mathbb{E}\left(\text{Tr}\left[(\mathbf{V}^T \mathbf{D}^{-1/2} \mathbf{V}(\mathbf{x}_{\text{approx}}^a - \mathbf{x}^a))(\mathbf{V}^T \mathbf{D}^{-1/2} \mathbf{V}(\mathbf{x}_{\text{approx}}^a - \mathbf{x}^a))^T\right]\right) \\
&= \mathbb{E}\|\mathbf{V}^T \mathbf{D}^{-1/2} \mathbf{V}(\mathbf{x}_{\text{approx}}^a - \mathbf{x}^a)\|_F^2
\end{aligned}$$

Therefore, $\mathbb{E}\|\mathbf{x}_{\text{approx}}^a - \mathbf{x}^a\|_{\mathbf{Q}^{-1}}^2$ measures the average total error in the posterior mean approximation normalized by the standard deviation of the error covariance \mathbf{Q} in the principal mode directions.

Adaptive Approximations An important consequence of Prop. 2.18 and Prop. 2.19 is that for a given rank k of the approximations, an optimal strategy can be devised to minimize the normalized error in the posterior error covariance and posterior mean. Indeed, based on Eq. (62)-(65) and Eq. (71)-(74), two regimes can be distinguished: if $\lambda_k < 1$, then $\mathbf{P}_{\mathbf{H}_{\Pi_{\text{dof}}}}^a$ and $\mathbf{x}_{\text{FR}_{\text{dof}}}^a$ should be chosen to minimize either the \mathbf{B} -normalized or the \mathbf{P}^a -normalized errors in \mathbf{P}^a and \mathbf{x}^a , respectively; if λ_k is significantly greater than 1, then a sensible strategy would be to use the updates $\mathbf{P}_{\Pi_{\text{dof}}}^a$ and $\mathbf{x}_{\Pi_{\text{dof}}}^a$ to approximate \mathbf{P}^a and \mathbf{x}^a , respectively. Note that this adaptive update procedure could also be used in the context of non-linear Gauss-Newton methods to improve the convergence rate of the minimization by using an optimal update for the quadratic solution at each linearization step (see Section 5.2).

3 Practical Implementation

3.1 Remarks on Eigendecompositions

The optimal approximations of the posterior error covariance matrix and the posterior mean described in Section 2.3 rely on the eigendecompositions of the large $n \times n$ matrices \mathbf{Q}_{var} and \mathbf{Q}_{dof} . In the high-dimensional framework considered in our study, those matrices cannot be formed explicitly, and therefore only matrix-free SVD algorithms can be employed (i.e., algorithms that use only matrix-vector products). In addition to its many theoretical benefits, including its interpretation as the solution of a projected Bayesian problem, the maximum-DOFS approximation associated with \mathbf{Q}_{dof} has important computational advantages through the simplification presented in Prop. 2.14. Indeed, the SVD of $\widehat{\mathbf{H}}_p$ does not involve direct inversions of large matrices¹. Assuming the tangent-linear and adjoint model are available, the SVD of $\widehat{\mathbf{H}}_p$ can be efficiently performed using

¹Although the observation error covariance matrix \mathbf{R} can be high-dimensional and non-diagonal, in practice covariance matrices and their inverses are constructed implicitly (e.g., Singh *et al.* (2011))

matrix-free algorithms such as Lanczos or randomized SVD methods (Lanczos, 1949; Halko *et al.*, 2011). The singular vectors \mathbf{v}_i computed from a truncated SVD of $\widehat{\mathbf{H}}_p$ can be used to obtain the singular vectors \mathbf{w}_i used to construct the approximated posterior mean $\mathbf{x}_{\Pi_{\text{dof}}}^a$ in Eq. (40), using the relation $\mathbf{w}_i = \mathbf{R}^{-1/2} \mathbf{H} \mathbf{B}^{1/2} \mathbf{v}_i$. Alternatively, a non-symmetric SVD algorithm such as that of Arnoldi (Golub and Van Loan, 2012) or Halko *et al.* (2011) (e.g., Alg. 5.1) can be used for direct computation of the SVD of the square-root of $\widehat{\mathbf{H}}_p$, $\widehat{\mathbf{H}}_p^{1/2} = \mathbf{R}^{-1/2} \mathbf{H} \mathbf{B}^{1/2}$. As shown in Prop. 2.15, computation of the singular vectors $\{\mathbf{w}_i\}$ can be avoided using Eq. (51). However, in the context of approximated SVD, one has to keep in mind that the equality between Eq. (40) and Eq. (51) does not strictly hold. Moreover, the singular vectors $\{\mathbf{w}_i\}$ can be useful for information content analysis, as discussed in Section 4.3. The possibility to efficiently compute the optimal approximations associated with the eigendecomposition of \mathbf{Q}_{dof} when both the control and the observation spaces are high-dimensional is in contrast with the optimal approximations associated with \mathbf{Q}_{var} (i.e., $\mathbf{x}_{\text{var}}^a$ and $\mathbf{P}_{\text{var}}^a$), since algebraic simplifications similar to Prop. 2.14 do not exist for \mathbf{Q}_{var} . In this case the $p \times p$ matrix of innovation statistics $\mathbf{H} \mathbf{B} \mathbf{H}^T + \mathbf{R}$ needs to be formed and inverted, and a matrix-free algorithm can then be used to compute the SVD of \mathbf{Q}_{var} (see Rem. 3.1).

Remarks 3.1. In the context of atmospheric source inversion, a typical case where the number of observations p is usually small enough to allow direct inversion of $\mathbf{H} \mathbf{B} \mathbf{H}^T + \mathbf{R}$ and compute the SVD of \mathbf{Q}_{var} is the inversion of (possibly high-dimensional) sources from a sparse network of *in situ* observations. In contrast, satellite-based inversions, for which p can be very large, may not allow $\mathbf{H} \mathbf{B} \mathbf{H}^T$ to be explicitly formed, unless the dataset is reduced prior to the inversion (e.g., using an aggregation scheme).

Remarks 3.2. In the case where the matrix of innovation statistics $\mathbf{H} \mathbf{B} \mathbf{H}^T + \mathbf{R}$ can be inverted explicitly, the full-dimensional analysis \mathbf{x}_a in Eq. (4) can be computed analytically even for control vectors with very large dimensions n (as long as the tangent-linear and adjoint models are available). However, even in that case, computing optimal approximations (based on either \mathbf{Q}_{var} or \mathbf{Q}_{dof}) is still useful in order to quantify the information content of the inversion, since the posterior error covariance and the model resolution matrices are both of dimension $n \times n$. To this aim, Eq. (41), (59), (60) and (42) can be used to efficiently extract subsets of elements (e.g., the entire diagonal) from the approximated posterior error covariance or model data resolution matrices.

3.2 Randomized Singular Value Decomposition

The most widely used matrix-free SVD algorithms are based on the Lanczos method, which computes the dominant eigenvectors and eigenvalues of an Hermitian matrix using Krylov subspace iterations (Golub and Van Loan, 2012). Recently, randomized SVD methods have attracted interest due to their proven accuracy and high scalability for a large variety of problems. In this Section, we describe a randomized SVD algorithm, some of its theoretical properties, as well as a practical probabilistic error estimate for the approximation. The use of this randomized SVD method allows critical improvement in computational performance that we shall exploit in a numerical experiment in the context of large-scale atmospheric source inversions (see Section 4).

3.2.1 Principle

Randomization algorithms are powerful and modern tools to perform matrix decomposition. Some of their key advantages compared to standard Krylov subspace methods are their inherent stability and the possibility to massively parallelize the computations. Recently, Halko *et al.* (2011) presented an extensive analysis of the theoretical and computational properties of randomized methods to compute approximate matrix decomposition, including low-rank SVDs. The approach relies on the ability to efficiently approximate the range of a matrix \mathbf{A} using a relatively small

sample of image vectors $\{\mathbf{y}^{(i)} = \mathbf{A}\boldsymbol{\omega}^{(i)}, i = 1, \dots, k\}$, where the input vectors $\boldsymbol{\omega}^{(i)}$ are independent vectors with i.i.d. Gaussian entries. The quality of the approximation for the range of \mathbf{A} can be objectively determined by evaluating the spectral norm of the difference between the original matrix and its projection onto the subspace defined by the random images, i.e., one wants:

$$\|(\mathbf{Id} - \mathbf{Q}\mathbf{Q}^T)\mathbf{A}\| \leq \epsilon, \quad (88)$$

where ϵ is some tolerance level, $\|\cdot\|$ is the spectral norm, and \mathbf{Q} is the matrix whose columns form an orthonormal basis of the subspace spanned by $\{\mathbf{y}^{(i)}, i = 1, \dots, k\}$. Once a satisfactory level of precision for the range has been reached, the SVD can be performed in the reduced space (defined by \mathbf{Q}) using dense matrix algebra, and the resulting singular vectors projected back onto the original space. Here we describe an algorithm especially adapted to the treatment of large Hermitian matrices involving expensive PDE solvers (in our case, the transport model \mathbf{H} and its adjoint \mathbf{H}^T). The reader is referred to Halko *et al.* (2011) for a complete review and explanation of those techniques in other contexts. The following algorithm allows one to compute an approximate truncated SVD of an Hermitian matrix using the randomized approach (Halko *et al.*, 2011). It uses only matrix-vector products, and is highly parallelizable.

Algorithm 1 One-Pass Eigenvalue Decomposition

Given \mathbf{A} a $n \times n$ Hermitian matrix and $\boldsymbol{\Omega}$ a $n \times k$ random Gaussian test matrix ($k \ll n$):

- 1: Form the $n \times k$ matrix $\mathbf{Y} = \mathbf{A}\boldsymbol{\Omega}$
 - 2: Construct a $n \times k$ matrix \mathbf{Q} whose columns form an orthonormal basis for the range of \mathbf{Y} , using, e.g., a QR or SVD factorization
 - 3: Form the $k \times k$ matrix $\mathbf{B} \equiv (\mathbf{Q}^T \mathbf{Y})(\mathbf{Q}^T \boldsymbol{\Omega})^{-1} (\approx \mathbf{Q}^T \mathbf{A} \mathbf{Q})$
 - 4: Perform the SVD $\mathbf{B} = \mathbf{Z}\boldsymbol{\Lambda}\mathbf{Z}^T$
 - 5: Form the $n \times k$ matrix column of singular vectors $\mathbf{V} = \mathbf{Q}\mathbf{Z}$
 - 6: One has the approximation $\mathbf{A} \approx \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^T$
-

In Algorithm 1, the products $\mathbf{A}\boldsymbol{\Omega}[:, i]$ in step 1 generating the columns $\mathbf{Y}[:, i]$ can be all performed in parallel, which renders this method highly scalable. In our case, the matrix-vector product $\mathbf{A}\boldsymbol{\Omega}[:, i]$ amounts to integrating a PDE solver, which requires intensive computations (e.g., $\mathbf{A} = \mathbf{B}^{1/2}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{B}^{1/2}$ for the maximum DOF approximation). Therefore, assuming $k \ll n$, the cost of the one-pass algorithm is largely dominated by step 1, the remaining steps involving dense linear algebra in small dimension.

In practice, the number of input samples k is increased until relation (88) is verified. The spectral norm of the estimation error is not directly evaluated (doing so would entail computing the SVD of a large $n \times n$ matrix, which is precisely what we want to approximate), but can be estimated a posteriori using an inexpensive probabilistic approach (see Section 3.2.2).

3.2.2 Error Analysis

The precision of the approximate SVD generated by Algorithm 1 depends on the error in the estimation of the range in (88), which itself depends (for a given number of samples) on the shape of the singular value spectra (i.e., fast or slow decay) and the dimension n . The following result demonstrates this dependence by providing a bound for the average spectral error (Halko *et al.*, 2011):

Proposition 3.1 (Average Spectral Error). *Let \mathbf{A} be a $n \times n$ matrix with eigenvalues $\sigma_1 \leq \dots \leq \sigma_n$. Let \mathbf{Q} be a $n \times (k+q)$ orthonormal basis that approximates the range of \mathbf{A} ($k+q \ll n$),*

generated from steps 1-2 of Algorithm 1. One has the following error bound:

$$\mathbb{E} \left(\|\mathbf{A} - \mathbf{Q}\mathbf{Q}^T \mathbf{A}\| \right) \leq \left[1 + \frac{4\sqrt{k+q}}{q-1} \sqrt{n} \right] \sigma_{k+1} \quad (89)$$

In Prop. 3.1, k is the targeted rank for the approximation, and l is an oversampling parameter. Note that the smallest possible spectral error for a rank- k approximation is σ_{k+1} [ref]. This formula shows that the average spectral error lies within a small polynomial factor of the theoretical minimum. In particular, one also sees that for very large n the bound is not significantly modified when n is increased (factor \sqrt{n}). Moreover, increasing the oversampling parameter q rapidly decreases the amplification factor $\frac{4\sqrt{k+q}}{q-1}$. As a result, in practice using an oversampling parameter $q \approx 5$ yields very good results.

In order to efficiently compute an estimate of the spectral error (88), one can use the following result, which provides a probabilistic bound based on sample error estimates available for free (Halko *et al.*, 2011):

Proposition 3.2 ((Cost-Free) Posterior Probabilistic Error Bound For Range Approximation). *Let \mathbf{A} be a $n \times n$ matrix. Let \mathbf{Q} be an orthonormal basis that approximates the range of \mathbf{A} , generated from steps 1-2 of Algorithm 1, and a set of l independent vectors with i.i.d. Gaussian entries $\{\boldsymbol{\omega}^{(i)}, i = 1, \dots, l\}$. One has the following error bound:*

$$\|\mathbf{A} - \mathbf{Q}\mathbf{Q}^T \mathbf{A}\| \leq 10 \sqrt{\frac{2}{\pi}} \max_{i=1, \dots, l} \|(\mathbf{A} - \mathbf{Q}\mathbf{Q}^T \mathbf{A})\boldsymbol{\omega}^{(i)}\|, \quad (90)$$

with a probability $1 - 10^{-l}$

Based on this result, one sees that using only two samples ($l = 2$) to estimate the right-hand side of (90) provides a bound on the spectral error with a probability 0.99. Note that samples of the form $\mathbf{A}\boldsymbol{\omega}^{(i)}$ are computed at step 1 of Alg. 1 to construct \mathbf{Q} . Therefore, a practical implementation of the a posteriori error bound estimate would be to use two samples out of the total set of samples generated at step 1 of Alg. 1 to derive the probabilistic bound in (90), while the remaining samples are used to compute \mathbf{Q} . An adaptive range finder method using similar principles to build a matrix \mathbf{Q} associated with a desired spectral error tolerance ϵ can also be found in (Halko *et al.*, 2011) (see Alg. 4.2).

4 Numerical Illustrations

In this Section we first illustrate the theoretical results obtained in Section 2 using a small inverse problem ($n = 300$). This enables exact computation of the SVD involved in the optimal approximations of the Bayesian problem, and direct evaluation of the performance of the associated posterior mean and posterior error covariance estimates against the true solutions. We then test the performance of the algorithm for a large-scale experiment using the randomized SVD method described in Section 3.2.1.

4.1 Atmospheric Source Inversion Problem

Our numerical experiments are carried out in the context of an atmospheric transport source inversion problem. The setup consists of an Observation Simulation System Experiment (OSSE) where pseudo-observations of methane columns (XCH_4) from a Short Wave Infrared (SWIR) instrument in low-earth orbit are generated and used to optimize randomly perturbed prior methane fluxes over North America. A nested domain with spatial resolution ($0.5^\circ \times 0.7^\circ$) is used and one scaling factor is optimized for each grid-cell for the month of July 2008, which corresponds to an initial

control space of dimension $n = 151 \times 121 = 18,271$. A uniform prior error standard deviation of 40% is assumed for the CH_4 fluxes throughout the domain, with no spatial error correlations (diagonal \mathbf{B} matrix), and the observational error standard deviations are uniformly set to 8 ppb, with no spatial or temporal correlations. More information about the general configuration of this type of OSSE experiment can be found in Bousserez *et al.* (2016). The atmospheric transport (forward model, H) is simulated using GEOS-Chem, which is an offline atmospheric transport model widely used in the atmospheric chemistry community. The model configuration we used is described in Wecht *et al.* (2014). The adjoint of GEOS-Chem, also employed in our experiment, is described in Henze *et al.* (2007) and has been extensively used in previous sensitivity and inverse modeling studies (Kopacz *et al.*, 2009; Jiang *et al.*, 2011; Xu *et al.*, 2013; Wells *et al.*, 2015). Figure 1 shows a map of the prior CH_4 emissions over North America. Emissions are geographically contrasting and highly variable. Since in our setup the prior error standard deviation is proportional to the prior emission magnitude, a similar high spatial variability is obtained for the prior error variances.

Note that the uniform diagonal \mathbf{B} matrix used in our setup implies that the matrices \mathbf{Q}_{dof} and \mathbf{Q}_{var} as well as their associated approximations will coincide modulo a scalar multiplication (this is obviously not the case for a general \mathbf{B} matrix). This simplified configuration allows a clear interpretation of the results in term of observational constraints (e.g., all posterior error correlations are due to the observations only), while demonstrating the theoretical properties and numerical efficiency of the optimal approximations. As discussed in Section 3.1, the approximations based on the solution of the maximum-DOFS projection have clear theoretical and practical advantages compared to the approximations derived from the eigendecomposition of \mathbf{Q}_{var} . Therefore, contrasting the characteristics of these two types of approximations in the general case (i.e., non-diagonal \mathbf{B}) was not viewed as a priority for our analysis.

4.2 Convergence Analysis for a Small Problem

The convergence properties of the optimal posterior mean and posterior error covariance approximations are investigated using a reduced version of the source inversion problem described in Section 4.1. In this experiment, the control vector dimension is reduced to $n = 300$ by selecting the model grid-cells which correspond to the first 300 highest gradients of the 4D-Var cost function (2) with respect to the emission scaling factors. With this setup, the Hessian of the cost function is explicitly calculated using the finite-difference method combined with adjoint model integrations. More specifically, the following formula is used to estimate each element of the Hessian matrix:

$$\hat{\mathbf{H}}_{i,j} \approx \frac{(\nabla J(\mathbf{x} + \epsilon_i) - \nabla J(\mathbf{x}))_j}{\epsilon}, \quad (91)$$

where $\epsilon_i = (\epsilon \delta_{i,k})_k$, δ is the Kronecker delta, and ϵ is a small real number. For this experiment $\epsilon = 0.01$, which corresponds to a 1% perturbation of the CH_4 flux for a particular grid cell. Since $\hat{\mathbf{H}}$ is symmetric, this calculation requires $n + 1$ gradient calculations, which corresponds to 301 forward model integrations and 301 adjoint model integrations for the reduced problem. The fact that these gradient calculations can be performed in parallel makes the computation efficient. The inverse of the Hessian matrix $\hat{\mathbf{H}}$ provides the posterior error covariance matrix \mathbf{P}^a , which is used to compute the exact posterior mean using formulas (3). The prior-preconditioned Hessian $\hat{\mathbf{H}}_p \equiv \mathbf{B}^{1/2} \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \mathbf{B}^{1/2}$ is also computed explicitly from the Hessian finite-difference estimate (91) using: $\hat{\mathbf{H}}_p = \mathbf{B}^{-1/2} (\hat{\mathbf{H}} - \mathbf{B}) \mathbf{B}^{-1/2}$.

Figure 2 shows the singular value spectra of the prior-preconditioned Hessian of the reduced inverse problem. The spectra shows a fast decrease of the first 20 singular values (by an order of magnitude), followed by a slow decrease. The basis for the rank- k maximum-DOFS projection is made of the first k singular vectors of $\hat{\mathbf{H}}_p$, which are computed exactly here. Moreover, according to Corollary 2.16, the DOFS of the inversion for a rank- k optimal projection is equal to $\sum_{i=1}^k \lambda_i / (1 +$

λ_i), where the $\{\lambda_i, i = 1, \dots, k\}$ are the first k singular values of $\hat{\mathbf{H}}_p$. The DOFS for our reduced inverse problem is ~ 43 . Figure 3 shows the \mathbf{P}^a -weighted error in the posterior error covariance matrix approximations $\mathbf{P}_{\mathbf{H}_{\Pi_{\text{dof}}}}^a$ and $\mathbf{P}_{\Pi_{\text{dof}}}^a$ defined by Eq. (63) and (65), respectively. For the sake of simplicity only the \mathbf{P}^a -weighted errors are considered here. As expected, for small ranks k (here, for $k < 3$) the low-rank approximation $\mathbf{P}_{\Pi_{\text{dof}}}^a$ is associated with a smaller error than the low-rank update approximation $\mathbf{P}_{\mathbf{H}_{\Pi_{\text{dof}}}}^a$. However, the approximation error associated with $\mathbf{P}_{\mathbf{H}_{\Pi_{\text{dof}}}}^a$ shows an exponential decrease and is about an order of magnitude smaller than the approximation error of $\mathbf{P}_{\Pi_{\text{dof}}}^a$ for $k > 50$.

Also shown on Fig. 3 is the relative error in the DOFS approximation for solution of the maximum-DOFS projection ($\mathbf{A}_{\Pi_{\text{dof}}}$), as well as the relative error in the total variance approximations $\text{Tr}(\mathbf{P}_{\mathbf{H}_{\Pi_{\text{dof}}}}^a)$ and $\text{Tr}(\mathbf{P}_{\Pi_{\text{dof}}}^a)$, as a function of the rank k of the approximation. The results for the DOFS show that more than 80% of the information content of the inversion is captured by the first 120 modes, with a fast decrease of the error for the first third of the spectra followed by a slower decrease. Interestingly, the low-rank update posterior error covariance approximation $\mathbf{P}_{\mathbf{H}_{\Pi_{\text{dof}}}}^a$ shows much better performances than the low-rank approximation $\mathbf{P}_{\Pi_{\text{dof}}}^a$, even for small values of the rank k . These results show that, for our experiment, the low-rank update approximation $\mathbf{P}_{\mathbf{H}_{\Pi_{\text{dof}}}}^a$ should be chosen when estimating posterior error variances. Overall, our numerical tests demonstrate the fast convergence of the low-rank update approximation $\mathbf{P}_{\mathbf{H}_{\Pi_{\text{dof}}}}^a$ for different information content metrics.

In addition to analyzing the convergence globally, it may be of interest to analyze the local behavior of this approximation. Figures 4 and 5 illustrate the convergence of the approximated error variances for $\mathbf{P}_{\mathbf{H}_{\Pi_{\text{dof}}}}^a$ for all control vector elements. Figure 4 represents the spatial distribution of the true posterior error variances as well as the approximated posterior error variances for the ranks $k = 40$, $k = 100$, and $k = 200$, while Fig. 5 shows the corresponding scatterplots and linear regression fits for each of those ranks. A very good accuracy of the approximated posterior error variances is observed for all ranks. This is further confirmed by the linear regression analysis, with a Pearson correlation coefficient of about 1 and an almost perfect regression line (1:1) for all ranks. From Fig. 5 it is evident that increasing the rank of the approximation above 40 does not significantly improve the results, which is consistent with the results obtained for the total error variance in Fig. 3.

Similarly to the posterior error covariance approximations, we now investigate the convergence properties of the posterior mean approximations described in Prop. 2.19. Again, for the sake of simplicity only the \mathbf{P}^a -weighted errors are considered here. Figure 6 shows the expectancy (or average) of the total \mathbf{P}^a -weighted error in the approximated posterior mean (or Bayes risk) for the solution of the maximum-DOFS projection $\mathbf{x}_{\Pi_{\text{dof}}}^a$ and for the full-rank posterior mean approximation $\mathbf{x}_{\text{FR}_{\text{dof}}}^a$, as a function of the rank. The results for one single realization of the prior and the observations are also shown. The Bayes risk for each rank k is calculated using Eq. (72) and (74), while the \mathbf{P}^a -weighted posterior mean error for one single realization is calculated by explicitly computing the error for one particular instance of the prior and observation probability distributions. As expected from the theory, for small values of the rank ($k < 10$) the error associated with the low-rank projection $\mathbf{x}_{\Pi_{\text{dof}}}^a$ is much smaller (by several order of magnitude) than the error associated with the full-rank approximation $\mathbf{x}_{\text{FR}_{\text{dof}}}^a$. However the full-rank approximation $\mathbf{x}_{\text{FR}_{\text{dof}}}^a$ becomes rapidly the most accurate (for $k > 10$), with an exponential decay of the error. The results for one realization of the prior and observation statistics also suggest very small deviations of the \mathbf{P}^a -weighted error from its mean behavior, which is consistent with previous findings in Spantini *et al.* (2015).

Similar to the posterior error analysis, it may be of interest to analyze locally the convergence of the approximated posterior mean, i.e., in our case, the spatial distribution of the posterior flux increments. Figure 7 shows the spatial distribution of the true and approximated posterior flux

increments for the solution of the maximum-DOFS projection $\mathbf{x}_{\Pi_{\text{dof}}}^a$ and the full-rank posterior increment approximation $\mathbf{x}_{\text{FR}_{\text{dof}}}^a$, for the ranks $k = 5$ and $k = 100$. The contrast between the low-rank nature of the maximum-DOFS solution $\mathbf{x}_{\Pi_{\text{dof}}}^a$ and the full-rank nature of $\mathbf{x}_{\text{FR}_{\text{dof}}}^a$ is evident for $k = 5$. As shown in Fig. 6, for a rank $k = 5$ the maximum-DOFS solution $\mathbf{x}_{\Pi_{\text{dof}}}^a$ is associated with a smaller \mathbf{P}^a -normalized error (Bayes risk) than $\mathbf{x}_{\text{FR}_{\text{dof}}}^a$. Although the full-rank approximation $\mathbf{x}_{\text{FR}_{\text{dof}}}^a$ better captures the true posterior increment distribution over large areas (e.g., Canada) compared to $\mathbf{x}_{\Pi_{\text{dof}}}^a$, those areas are associated with large posterior errors (see Figure 4), and thus are attributed less weight in the \mathbf{P}^a -normalized posterior mean score than regions over the east of the US domain associated with small posterior errors. The better performance of the low-rank maximum-DOFS solution $\mathbf{x}_{\Pi_{\text{dof}}}^a$ over those regions with small posterior errors explains its overall better score. Consistent with our previous analysis (see Fig. 6), for $k = 100$, the full-rank posterior increment approximation $\mathbf{x}_{\text{FR}_{\text{dof}}}^a$ better reproduces the spatial distribution of the true posterior increment across the whole domain, with now similar performances as the maximum-DOFS posterior increment $\mathbf{x}_{\Pi_{\text{dof}}}^a$ over regions associated with small posterior errors (see, e.g., the eastern US).

4.3 Performance for a Large-Scale Experiment

In this Section we illustrate the efficiency of combining the optimal approximation methods with the randomized SVD algorithm by applying this approach to the full-dimensional source inversion problem defined in Section 4.1. Since the dimension of the control vector is now $n = 151 \times 121 = 18,271$, explicitly forming the prior-preconditioned Hessian matrix and computing its SVD using dense linear algebra is not practical. Therefore, we use the One-Pass SVD algorithm described in Alg. 1 to compute the eigenvectors and eigenvalues of $\hat{\mathbf{H}}_p$. The computation can be performed in parallel, since the k matrix-vector products of $\hat{\mathbf{H}}_p$ with the columns of the $n \times k$ random Gaussian test matrix $\mathbf{\Omega}$ can be performed independently. In our context, the computational cost of the method is largely dominated by the Hessian matrix-vector products used to build the basis for the reduced space \mathbf{Q} (see step 1 of Alg. 1), since each of them amounts to integrating one tangent linear and adjoint model. The shape of the approximated eigenvalue spectra of $\hat{\mathbf{H}}_p$, not shown here, is comparable to the one obtained with the small problem of Section 4.2, with an exponential decay followed by a long flat tail, typical of severely underconstrained inverse problems.

Figure 8 shows the probabilistic spectral error bound calculated using Prop. 3.2, as a function of the number of samples (i.e., the number of columns k of $\mathbf{\Omega}$) used in the randomized SVD estimate. A fast decrease of the spectral error bound is observed for the first 100 samples, followed by a smaller decrease between 100 and 400 samples. It appears clearly that increasing the number of samples beyond 200 does not significantly reduce the spectral error bound (~ 3), which suggests a reduced basis of 200 vectors would provide a reasonably good approximation of the range of $\hat{\mathbf{H}}_p$ for our problem. Here we used those 400 samples to form an orthonormal basis \mathbf{Q} of the subspace in which the low-rank SVD was performed. We note that no significant differences were found in the approximated posterior error covariances or the approximated posterior mean updates between computations using 200 or 400 samples.

The analytical expressions for the posterior error covariance approximations, $\mathbf{P}_{\mathbf{H}_{\Pi_{\text{dof}}}}^a$ (Eq. (59)) and $\mathbf{P}_{\Pi_{\text{dof}}}^a$ (Eq. (49)), and for the maximum-DOFS model resolution matrix approximation, $\mathbf{A}_{\Pi_{\text{dof}}}$ (Eq. (50)), can be used to efficiently compute any matrix-vector product involving those matrices, and in particular to extract subsets of their elements. As an example, Fig. 9 shows the diagonal of the model resolution matrix $\mathbf{A}_{\Pi_{\text{dof}}}$ of the rank-400 maximum-DOFs projection, which represents the observational constraints for each flux estimate. More specifically, each value on the diagonal quantifies the relative contribution of the observations to the total information content, with respect to the prior information. As discussed previously, it also corresponds to the sensitivity of the posterior mean flux to its true value (see Eq.(7)). The sum of the diagonal elements of $\mathbf{A}_{\Pi_{\text{dof}}}$ is

the DOFS for the projected problem, which is equal to 46. Therefore, 46 independent pieces of information (i.e., flux modes) can be obtained from the observational constraints. As explained in Section 2.2.4, the eigenvectors of the model resolution matrix represent precisely the modes that are independently constrained by the observations (with respect to the prior information), and are given by $\{\mathbf{B}^{1/2}\mathbf{v}_i, i = 1, \dots, k\}$, where the $\{\mathbf{v}_i, i = 1, \dots, k\}$ are the first k eigenvectors of the prior-preconditioned Hessian $\hat{\mathbf{H}}_p$. Figure 10 shows the 1st, 2nd and 5th eigenvectors of $\hat{\mathbf{H}}_p$, which in the case of a uniform diagonal \mathbf{B} matrix are therefore scalar multiples of the eigenvectors of the model resolution matrix $\mathbf{A}_{\Pi_{\text{dof}}}$. Additionally, the corresponding modes in observation space, that is, the left singular vectors $\{\mathbf{w}_i, i = 1, \dots, k\}$ of the square-root of $\hat{\mathbf{H}}_p$, are also shown and were computed by propagating each singular vectors \mathbf{v}_i using the relation $\mathbf{w}_i = \mathbf{R}^{-1/2}\mathbf{H}^T\mathbf{B}^{1/2}\mathbf{v}_i$ (see Section 3.1). As shown, the relative contribution of the observations to the information content of each posterior mode is greater than 80% for all three modes. The three modes correspond to clearly distinct geographical patterns. The results show that the satellite observations allow one to constrain CH_4 emissions at high (almost grid-scale) spatial resolution over the Toronto (1st mode) and (to a lesser extent) the Los Angeles areas, but that constraints between the New York and Appalachian regions tend to be correlated. These figures illustrate the potential of the maximum-DOFS low-rank projection problem as a robust framework to objectively characterize the information content of an inversion.

In addition to providing useful tools to analyze fundamental properties of the inverse problem, the joint use of the optimal approximations with the randomized SVD approach provides a fast method to compute the posterior solution. The efficiency of the approach can be illustrated by comparing the number of iterations required for convergence of a standard iterative minimization routine with the number of samples needed for the randomized SVD to provide similar results. Figure 11 shows the posterior scaling factor increments obtained from a BFGS minimization with 40 iterations and from our adaptive posterior mean approximation method using randomized SVD computations with 200 samples. In this case, since $\lambda_{200} < 1$, the approximated posterior mean chosen is $\mathbf{x}_{\text{FR}_{\text{dof}}}^a$ (see Section 2.3.2). The results show that a parallel implementation of the randomized SVD approach using only 200 samples provides posterior flux increments comparable to 40 iterations of the BFGS algorithm over most areas (with differences rarely exceeding 0.10). The moderate number of samples for this test allowed us to run all independent simulations in the randomized SVD algorithm simultaneously, resulting in a wall time computation of ~ 72 mins. Comparatively, the BFGS minimization algorithm required a wall time 40 times longer than the randomized SVD, with a total of 48 hours. These results clearly illustrate the benefit of using the randomized SVD approach to drastically reduce the computational cost of large-scale inverse problems.

As discussed in Halko *et al.* (2011), the efficiency of randomized range approximation methods is driven by the rate of decrease of the singular value spectra, and is therefore problem-dependent. As suggested by Thm. 3.1, a sharp decrease of the singular values is associated with a more accurate estimate of the range of the matrix. We note that many inverse problems in geophysics, including atmospheric source inversion and weather data assimilation problems, generally exhibit this desirable behavior. However, for inverse problems whose singular values decay slowly, randomized power-iteration methods can still be used, which provides much more accurate results while mitigating the cost of the SVD compared to standard Krylov subspace methods. Their efficiency stems from exploiting parallel implementation of randomized estimates combined with a small number of power-iterations (Halko *et al.*, 2011).

5 Link With Methods in Data Assimilation

In this Section we discuss some interesting links between the theory developed in this paper and methods used in operational data assimilation centers. In Section 5.1, we recall the incremental

4D-Var algorithm and its different square-root formulations, and explain how it is closely related to the maximum-DOFS low-rank projection. An overlooked issue related to the representativeness error when singular square-roots are used is also discussed. In Section 5.2, we propose an improved implementation of the incremental 4D-Var method, combining the optimality results of Section 2.3 with the randomized SVD technique described in Section 3.2.

5.1 Incremental 4D-Var

5.1.1 Principle

Most large-scale inverse problems, such as those encountered in atmospheric data assimilation, are solved by minimizing the least-squares cost function (2). Operational DA centers often use the incremental 4D-Var technique (e.g., Courtier *et al.* (1994)), which consists of a sequence of quadratic CG minimizations based on linearizations of the forward model. Each quadratic minimization is called an inner-loop, and is often performed using simplified tangent-linear and adjoint models. After each quadratic minimization, the updated posterior mean is propagated through the non-linear model H and compared to the observations. This step, called the outer-loop, is repeated until a convergence criterion is reached.

In its original form, the linear (or quadratic) least-squares problem can be severely ill-conditioned, i.e., the gap between the eigenvalues of the Hessian of the cost function can be very large. This can prevent fast convergence of the CG algorithm. One widely used remedy is preconditioning, which consists of solving the least-squares problem in a basis where the eigenvalues of the Hessian of the cost function are as tightly clustered as possible. In the ideal case, where the Hessian is the identity matrix, the convergence is obtained in one iteration, which is why the matrix of change of basis (or preconditioner) should be constructed so as to best approximate a square-root of the Hessian. In practice, only limited prior knowledge of the Hessian is available (note that knowing this matrix perfectly amounts to solving the least-squares problem). Since the Hessian can be written as a positive low-rank update to the inverse prior error covariance matrix ($\nabla^2 J(\mathbf{x}^a) = \mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}$), a common approach is to use the square-root of \mathbf{B} as a preconditioner, which one shall note \mathbf{L} , such as $\mathbf{B} = \mathbf{L}\mathbf{L}^T$. This change of variable leads to the following least-squares problem for each inner-loop of the incremental 4D-Var algorithm:

$$\begin{aligned} \min_{\mathbf{z}} \tilde{J}(\mathbf{z}) &= \frac{1}{2}(\mathbf{d} - \mathbf{H}\mathbf{L}\mathbf{z})^T \mathbf{R}^{-1}(\mathbf{d} - \mathbf{H}\mathbf{L}\mathbf{z}) + \frac{1}{2}\mathbf{z}^T \mathbf{z} \\ \mathbf{B} &= \mathbf{L}\mathbf{L}^T, \\ \mathbf{x} &= \mathbf{L}\mathbf{z}, \end{aligned} \tag{92}$$

where \mathbf{x} is the increment and \mathbf{d} the innovation.

5.1.2 Square-Root Formulations

There are two main types of square-root formulations used in current 4D-Var DA systems. The Control Variable Transform (CVT) consists of modeling the prior error covariance matrix \mathbf{B} implicitly, by projecting the control vector onto a basis where its elements are uncorrelated (i.e., their covariance matrix is the identity matrix) (Bannister, 2008). In this approach, the change of basis is constructed using operators that impose, e.g., dynamical balances, or filtering out modes of variability that one does not wish to include in the posterior update. In addition to allowing an implicit representation of a matrix that would otherwise be too large to store in computer memory ($n \sim 10^8$ in operational NWP), this technique provides an efficient preconditioning of the variational minimization (Courtier *et al.*, 1994). With the CVT technique, the product of all the operators that decorrelate the control variables forms the square-root \mathbf{L} .

The other approach to the square-root formulation is the ensemble-based method. An ensemble-based square-root consists of modeling the matrix \mathbf{B} using an ensemble of forward model perturbation trajectories centered around their mean, i.e.:

$$\mathbf{B}_{\text{ens}} = \mathbf{W}\mathbf{W}^T, \quad \mathbf{W} = \frac{1}{\sqrt{K-1}}(\mathbf{w}_1 - \bar{\mathbf{w}}, \dots, \mathbf{w}_K - \bar{\mathbf{w}}), \quad (93)$$

where $(\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_K)$ represents an ensemble of K perturbed states and $\bar{\mathbf{w}}$ the mean of the ensemble. This approach is used, e.g., in ensemble-based DA methods such as the Ensemble Adjustment Kalman filter (EAKF) (Anderson, 2001), or the Ensemble-Variational (EnVar) algorithm (Lorenc, 2003). Localization techniques are usually required to mitigate undersampling noise in the approximated error variances and error correlations (e.g., Ménétrier *et al.* (2015)). In practice, it may be desirable to implement a hybrid formulation, in which \mathbf{B} is a weighted average of a static full-rank matrix \mathbf{B}_{stat} that represents, e.g., a climatology, and an ensemble-based flow-dependent matrix \mathbf{B}_{ens} , so that \mathbf{B} is eventually modeled as:

$$\mathbf{B}_{\text{hyb}} = (1 - \beta)\mathbf{B}_{\text{stat}} + \beta\mathbf{B}_{\text{ens}} \circ \mathbf{C}, \quad (94)$$

where \mathbf{C} is made of compactly supported and space-limited covariance functions (Gaspari and Cohn, 1999), \circ represents the Schur product between two matrices, and β is a scalar verifying $0 < \beta < 1$.

The matrix \mathbf{B}_{hyb} can be defined only at the initial time, t_0 , of the variational window, which is the En4DVar method, or throughout the assimilation window (4D matrix), which is the 4DEnVar (or EnVar) approach (Desroziers *et al.*, 2014; Lorenc, 2003). Based on this very general square-root formulation of incremental 4D-Var, we will now establish some theoretical links between optimizations performed in current operational DA systems and the optimal low-rank approximation approaches presented in this study.

5.1.3 Preconditioned Conjugate-Gradient as an Optimal Low-Rank Projection

The following proposition establishes the theoretical equivalence (i.e., modulo approximation errors) between the posterior mean of the maximum-DOFS rank- k projection and the solution obtained after k iteration of the preconditioned CG algorithm. It stems from the close relationship between the CG algorithm and the Lanczos eigenvalue decomposition (Meurant and Strakoš, 2006).

Proposition 5.1 (Optimality of Preconditioned Conjugate-Gradient Minimizations). *Let us consider the \mathbf{L} -preconditioned least-squares minimization problem (92), where $\mathbf{B} = \mathbf{L}\mathbf{L}^T$ and $\text{rank}(\mathbf{L}) = \text{rank}(\mathbf{B}) = n$. One notes $\mathbf{x}_{\Pi_{\text{dof}}}^a$ the posterior mean of the rank- k maximum-DOFS projection (40) and \mathbf{z}_k^a the solution obtained after k (non-converged) iterations of a conjugate-gradient minimization starting from $\mathbf{z}_0^a = 0$. One has:*

$$\mathbf{L}\mathbf{z}_k^a = \mathbf{x}_{\Pi_{\text{dof}}}^a \quad (95)$$

Proof. We first rewrite the cost function in (92):

$$\begin{aligned} \tilde{J}(\mathbf{z}) &= \frac{1}{2}(\mathbf{d} - \mathbf{H}\mathbf{L}\mathbf{z})^T \mathbf{R}^{-1}(\mathbf{d} - \mathbf{H}\mathbf{L}\mathbf{z}) + \frac{1}{2}\mathbf{z}^T \mathbf{z} \\ &= \frac{1}{2}[\mathbf{z}^T (\mathbf{L}^T \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \mathbf{L} + \mathbf{Id}) \mathbf{z} - \mathbf{z}^T \mathbf{L}^T \mathbf{H}^T \mathbf{R}^{-1} \mathbf{d} - \mathbf{d}^T \mathbf{R}^{-1} \mathbf{H} \mathbf{L} \mathbf{z} + \mathbf{d}^T \mathbf{R}^{-1} \mathbf{d}] \\ &= \frac{1}{2}[\mathbf{z}^T \mathbf{A} \mathbf{z} - \mathbf{k}^T \mathbf{z} - \mathbf{z}^T \mathbf{k}] + \text{constant}, \end{aligned}$$

with $\mathbf{A} = \mathbf{L}^T \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \mathbf{L} + \mathbf{Id}$ and $\mathbf{b} = \mathbf{L}^T \mathbf{H}^T \mathbf{R}^{-1} \mathbf{d}$. Minimizing $\tilde{J}(\mathbf{z})$ is therefore equivalent to solving:

$$\mathbf{A} \mathbf{z} = \mathbf{b}$$

Starting with $\mathbf{z}_1 = \mathbf{0}$, the CG algorithm at iteration k produces (e.g., Golub and Van Loan (2012)):

$$\mathbf{z}_k^a = \operatorname{argmin}_{\mathbf{z} \in \mathcal{K}_k(\mathbf{A}, \mathbf{z}_1)} \|\mathbf{A}\mathbf{z} - \mathbf{b}\|^2, \quad (96)$$

where $\mathcal{K}_k(\mathbf{A}, \mathbf{z}_1)$ is the k -dimensional Krylov subspace associated with \mathbf{A} and initialized with the vector \mathbf{z}_1 , i.e., $\mathcal{K}_k(\mathbf{A}, \mathbf{z}_1) = \{\mathbf{z}_1, \mathbf{A}\mathbf{z}_1, \dots, \mathbf{A}^k \mathbf{z}_1\}$. One can rewrite (96):

$$\mathbf{z}_k^a = \operatorname{argmin}_{\mathbf{z}} \|\mathbf{A}\mathbf{Q}\mathbf{Q}^T \mathbf{z} - \mathbf{b}\|^2, \quad (97)$$

where \mathbf{Q} is a matrix whose columns form an orthonormal basis of $\mathcal{K}_k(\mathbf{A}, \mathbf{z}_1)$. Noting that $\mathcal{K}_k(\mathbf{A}, \mathbf{p}_1) \approx \operatorname{Im}(\{\mathbf{v}_i, i = 1, \dots, k\})$, where \mathbf{v}_i represents the i -th eigenvector of $\widehat{\mathbf{H}}_p$, which is a basic property of the CG algorithm and is related to its link with the Lanczos method (Golub and Van Loan, 2012). This is equivalent to assuming $\mathbf{Q}\mathbf{Q}^T \approx \mathbf{V}_k \mathbf{V}_k^T$. We therefore obtain the least-squares solution:

$$\begin{aligned} \mathbf{z}_k^a &= [\mathbf{A}\mathbf{Q}\mathbf{Q}^T]^+ \mathbf{b} \\ &= [\mathbf{A}\mathbf{V}_k \mathbf{V}_k^T]^+ \mathbf{b} \\ &= \mathbf{L} \left[\sum_{i=0}^k \mathbf{v}_i \mathbf{v}_i^T (1 - \lambda_i (1 + \lambda_i)^{-1}) \right] \mathbf{V} \mathbf{\Lambda}^{1/2} \mathbf{W} \mathbf{R}^{-1/2} \mathbf{d} \\ &= \mathbf{L} \left[\sum_{i=1}^k \lambda_i^{1/2} (1 + \lambda_i)^{-1} \mathbf{v}_i \mathbf{w}_i^T \right] \mathbf{R}^{-1/2} \mathbf{d} \\ &= \mathbf{x}_{\Pi_{\text{dof}}}^a, \end{aligned}$$

where we used $\mathbf{A} = \mathbf{V}(\mathbf{Id} - \mathbf{\Lambda}^{1/2}(\mathbf{Id} + \mathbf{\Lambda})^{-1})\mathbf{V}^T$ and $\mathbf{b} = \mathbf{V}\mathbf{\Lambda}^{1/2}\mathbf{W}\mathbf{R}^{-1/2}\mathbf{d}$. □

From Proposition 5.1, we see that the solution of the preconditioned CG minimization after k non-converging iterations is also the posterior mean of an optimal projection, in the sense of Prop. 2.19. Moreover, it is also the posterior mean solution of the projected Bayesian problem with maximum DOF. This latter fact is useful in term of interpretation, since it means that the non-converged solution (for the initial problem) is still the exact solution of a low-rank Bayesian problem, with an information content explicitly defined by Eq. (42). Preconditioning the CG algorithm with a square-root of the prior covariance \mathbf{B} was adopted for practical reasons in operational DA systems, that is, to improve the convergence of the minimization and to efficiently represent (implicitly) a high-dimensional \mathbf{B} matrix. Interestingly, our result provides also a theoretical justification for preconditioning the CG algorithm with a square-root of \mathbf{B} in quadratic 4D-Var minimizations. We also note that potentially better approximations of the posterior mean (i.e., $\mathbf{x}_{\text{FRdof}}^a$, see Fig. 6) are available.

5.1.4 Singular Square-Root Formulations and Representativeness Error

In the context of Gaussian pdf assumptions and maximum likelihood estimation, the use of a non-singular \mathbf{B} matrix is required, as evident from the presence of \mathbf{B}^{-1} in the cost function (2). The positive-definiteness of \mathbf{B} indicates that none of the control space directions are associated with a null probability density in the case of a Gaussian pdf. Square-root formulations of \mathbf{B} currently used in operational DA systems (CVT and/or ensemble-based) can violate this assumption. As emphasized recently by Ménétrier and Auligné (2015), in some DA systems the balance operators used to construct the CVT lead to a non-square square-root \mathbf{L} , and therefore implicitly model a singular \mathbf{B} matrix. Moreover, as previously mentioned, the number of trajectory perturbations, K , used to represent \mathbf{B} in ensemble-based approaches is very small compared to the dimension of the control vector, i.e., $K \ll n$, which also leads to singular square-roots for \mathbf{B} .

We note that preconditioned 4D-Var techniques that use a singular square-root for \mathbf{L} can be readily interpreted within the theoretical framework developed in Section 2.2. Indeed, solving the preconditioned quadratic 4D-Var minimization problem (92) with a singular square-root \mathbf{L} amounts to applying a two-step low-rank projection method, as described in Section 2.1.1, with $\mathbf{\Gamma} = (\mathbf{L}^T \mathbf{L})^{-1} \mathbf{L}^T$ and $\mathbf{\Gamma}^* = \mathbf{L}$. In this case, it is also easy to check that the prolongation operator verifies $\mathbf{\Gamma}^* = \mathbf{B} \mathbf{\Gamma}^T (\mathbf{\Gamma}^T \mathbf{B} \mathbf{\Gamma})^{-1}$, which makes it an optimal prolongation (see Thm. 2.6). Therefore, by Corollary 2.10, the exact solution of this singular preconditioned minimization problem (92) is the projection of the solution of the initial (full-rank) Bayesian problem, i.e., one has:

$$\mathbf{L} \mathbf{z}^a = \mathbf{\Pi}_L \mathbf{x}^a, \quad (98)$$

with $\mathbf{\Pi}_L = \mathbf{\Gamma}^* \mathbf{\Gamma}$. Note that (98) is verified if the matrix $\mathbf{R}_{\mathbf{\Pi}_L}$, which accounts for the representativeness error due to the restriction of the problem to the subspace spanned by the column of \mathbf{L} , is used in (92) instead of the initial observational error covariance matrix \mathbf{R} (i.e., (98) is the solution of the Bayesian problem $\mathcal{B}_{\mathbf{\Pi}_L} = (E, F, \mathbf{H} \mathbf{\Pi}_L, \mathbf{B}, \mathbf{R}_{\mathbf{\Pi}_L})$). It is worthwhile to note that when singular square-roots formulations are used in operational DA systems, the initial observational error covariance (\mathbf{R}) is not modified, which can theoretically lead to errors in the optimization. Since the representativeness error 2.2 cannot be computed in practice, it is useful to understand under which conditions it vanishes, which is the object of the following Proposition:

Proposition 5.2. *Let us consider \mathbf{L} a singular square-root of \mathbf{B} ($\text{rank}(\mathbf{L}) < n$), and $\mathbf{\Pi}_L = \mathbf{L}(\mathbf{L}^T \mathbf{L})^{-1} \mathbf{L}^T$ and $\mathbf{R}_{\mathbf{\Pi}_L}$ its associated projection and representativeness error, respectively. One has:*

$$\mathbf{R}_{\mathbf{\Pi}_L} = \mathbf{R} \Leftrightarrow \text{Im}(\mathbf{L})^\perp \subset \ker(\mathbf{H}) \Leftrightarrow \text{Im}(\mathbf{L})^\perp \subset \text{Im}(\mathbf{H}^T)^\perp, \quad (99)$$

where $^\perp$ denotes the orthogonal complement.

Proof. The second equivalence on the right is immediate since $\ker(\mathbf{H}) = \text{Im}(\mathbf{H}^T)^\perp$. To show the first equivalence, let us write explicitly the representativeness error:

$$\mathbf{R}_{\mathbf{\Pi}_L} = \mathbf{R} + \mathbf{H}(\mathbf{B} + \mathbf{\Pi}_L \mathbf{B} \mathbf{\Pi}_L - \mathbf{\Pi}_L \mathbf{B} - \mathbf{B} \mathbf{\Pi}_L) \mathbf{H}^T \quad (100)$$

Let us decompose the (full-rank) \mathbf{B} matrix as:

$$\mathbf{B} = \mathbf{B}_L + \mathbf{B}_{L^\perp} = \mathbf{Z} \mathbf{\Theta} \mathbf{Z}^T + \mathbf{Z}_\perp \mathbf{\Omega} \mathbf{Z}_\perp^T, \quad (101)$$

where $\mathbf{B}_L = \mathbf{L} \mathbf{L}^T = \mathbf{Z} \mathbf{\Theta} \mathbf{Z}^T$ is an eigendecomposition of \mathbf{B} in the subspace spanned by the columns of \mathbf{L} , and $\mathbf{B}_\perp = \mathbf{Z}_\perp \mathbf{\Omega} \mathbf{Z}_\perp^T$ is the expression of \mathbf{B} in a basis $\{\mathbf{z}_i^\perp, i = 1, \dots, k\}$ of the orthogonal complement of $\text{Im}(\{\mathbf{z}_i, i = 1, \dots, k\})$ (i.e., $\mathbf{z}_i^T \mathbf{z}_j^\perp = 0, \forall i, j$). Using (101) and the orthogonality properties, it is clear that $\mathbf{\Pi}_L \mathbf{B} = \mathbf{L} \mathbf{L}^T$. Therefore, one has: $\mathbf{B} + \mathbf{\Pi}_L \mathbf{B} \mathbf{\Pi}_L - \mathbf{\Pi}_L \mathbf{B} - \mathbf{B} \mathbf{\Pi}_L = \mathbf{B} - \mathbf{B} \mathbf{\Pi}_L = \mathbf{Z}_\perp \mathbf{\Omega} \mathbf{Z}_\perp^T$. Replacing this expression in (100), one obtains:

$$\mathbf{R}_{\mathbf{\Pi}_L} = \mathbf{R} + \mathbf{H} \mathbf{Z}_\perp \mathbf{\Omega} \mathbf{Z}_\perp^T \mathbf{H}^T \quad (102)$$

Using a square-root of $\mathbf{\Omega}^2$, one can also write:

$$\mathbf{R}_{\mathbf{\Pi}_L} = \mathbf{R} + \mathbf{M} \mathbf{M}^T, \quad (103)$$

² Note that $\mathbf{\Omega} = \mathbf{\Omega}^T$, from $\mathbf{Z}_\perp \mathbf{\Omega} \mathbf{Z}_\perp^T = \mathbf{Z}_\perp \mathbf{\Omega}^T \mathbf{Z}_\perp^T$ and multiplication by $(\mathbf{Z}_\perp^T \mathbf{Z}_\perp)^{-1} \mathbf{Z}_\perp^T$ and $\mathbf{Z}_\perp (\mathbf{Z}_\perp^T \mathbf{Z}_\perp)^{-1}$ on the left and right, respectively, which guarantees the existence of a square-root for $\mathbf{\Omega}$.

where $\mathbf{M} = \mathbf{H}\mathbf{Z}_\perp\boldsymbol{\Omega}^{1/2}\mathbf{Z}_\perp^T$. Therefore, one obtains the equivalence:

$$\mathbf{R}_{\Pi_L} = \mathbf{R} \Leftrightarrow \mathbf{M}\mathbf{M}^T = \mathbf{0} \quad (104)$$

$$\Leftrightarrow \mathbf{M} = \mathbf{0} \quad (105)$$

$$\Leftrightarrow \text{Im}(\mathbf{Z}_\perp\boldsymbol{\Omega}^{1/2}\mathbf{Z}_\perp^T) \subset \ker(\mathbf{H}) \quad (106)$$

$$\Leftrightarrow \text{Im}(\mathbf{Z}_\perp) \subset \ker(\mathbf{H}) \quad (107)$$

$$\Leftrightarrow \text{Im}(\mathbf{L})^\perp \subset \ker(\mathbf{H}), \quad (108)$$

where we used the fact that $\text{Im}(\mathbf{Z}_\perp\boldsymbol{\Omega}^{1/2}\mathbf{Z}_\perp^T) = \text{Im}(\mathbf{Z}_\perp)$ (since $\boldsymbol{\Omega}$ is full-rank) and $\text{Im}(\mathbf{Z}_\perp) = \text{Im}(\mathbf{L})^\perp$. \square

Diagnosing the Representativeness Error In practice, the right equivalence in (99) can be useful to diagnose the representativeness error. Indeed, one has the following necessary condition:

$$\mathbf{R}_{\Pi_L} = \mathbf{R} \Leftrightarrow \text{Im}(\mathbf{L})^\perp \subset \text{Im}(\mathbf{H}^T)^\perp \Rightarrow \text{Im}(\mathbf{H}^T) \subset \text{Im}(\mathbf{L}), \quad (109)$$

where we used the fact that $\text{Im}(\mathbf{A}) \subset \text{Im}(\mathbf{B}) \Rightarrow \text{Im}(\mathbf{B})^\perp \subset \text{Im}(\mathbf{A})^\perp$ for any two matrices \mathbf{A} and \mathbf{B} . The necessary condition on the right of (109) means that the subspace corresponding to non-zero sensitivities of the observations to the control space should be restricted to the range of the singular square-root of \mathbf{L} . We note that an approximation of both $\text{Im}(\mathbf{L})$ and $\text{Im}(\mathbf{H}^T)$ can be obtained using the randomized range finder described in Alg. 1 (step 1 and 2) for \mathbf{L} and \mathbf{H}^T , respectively. If one notes $\mathbf{Q}_{\mathbf{H}^T}$ an orthonormal basis for the range of \mathbf{H}^T and \mathbf{Q}_L an orthonormal basis for the range of \mathbf{L} , one has: $\text{Im}(\mathbf{H}^T) \subset \text{Im}(\mathbf{L}) \Leftrightarrow \mathbf{Q}_L\mathbf{Q}_L^T\mathbf{Q}_{\mathbf{H}^T} = \mathbf{Q}_{\mathbf{H}^T}$. Therefore, in the case where the orthonormal basis \mathbf{Q}_L and $\mathbf{Q}_{\mathbf{H}^T}$ are only approximations, a diagnostic for the representativeness error can consist of checking that:

$$\mathbf{Q}_L\mathbf{Q}_L^T\mathbf{Q}_{\mathbf{H}^T} \approx \mathbf{Q}_{\mathbf{H}^T} \Rightarrow \|(\text{Id} - \mathbf{Q}_L\mathbf{Q}_L^T)\mathbf{Q}_{\mathbf{H}^T}\|^2 \approx 0 \quad (110)$$

In a cycling DA context, $\text{Im}(\mathbf{H}^T)$ would need to be estimated at each DA cycle, while $\text{Im}(\mathbf{L})$ would be estimated only once. However, in Section 5.2 we discuss a strategy that would allow estimation of $\text{Im}(\mathbf{H}^T)$ as a by-product of the minimization of (92), where the CG algorithm is replaced by a randomized SVD approach.

Remarks 5.1. A recent study by Ménétrier and Auligné (2015) discusses the impact of singular square-root preconditionings in variational minimization. The discussion focuses on the idea that the change of variable $\mathbf{x} = \mathbf{L}\mathbf{v}$ should rigorously lead to a preconditioned background term of the cost function of the form $J_b(\mathbf{v}) = \mathbf{v}^T\mathbf{L}^T\mathbf{B}^{-1}\mathbf{L}\mathbf{v}$, which can be different from the formulation $J_b(\mathbf{v}) = \mathbf{v}^T\mathbf{v}$ always used in practice. The authors demonstrate that in the subspace where the CG minimization operates, the two background term formulations are equivalent. However, as described in the present paper (see Prop. 2.5), and since the singular preconditioning can be interpreted as an effective dimension reduction, the correct approach here is to consider $\mathbf{B}_\omega = \mathbf{L}^T\mathbf{B}\mathbf{L} = \text{Id}_m$ ($m < n$) as the prior error covariance matrix of the preconditioned cost function, since it is precisely the prior error covariance of the reduced Bayesian problem $\mathcal{B}_\omega = (E_\omega, F, \mathbf{H}\boldsymbol{\Gamma}^*, \boldsymbol{\Gamma}^T\mathbf{B}\boldsymbol{\Gamma}, \mathbf{R}_\Pi)$ (with $\boldsymbol{\Gamma} = (\mathbf{L}^T\mathbf{L})^{-1}\mathbf{L}^T$ and $\boldsymbol{\Gamma}^* = \boldsymbol{\Gamma}$). Therefore, one sees that the claim that there would be any potential inconsistency in using $J_b(\mathbf{v}) = \mathbf{v}^T\mathbf{v}$ in the preconditioned 4D-Var minimization has no theoretical basis. The inconsistency can only arise from neglecting the representativeness error, as explained in this Section.

5.2 Improving Incremental 4D-Var

In this Section we propose some improvements to the standard 4D-Var algorithm, leveraging the results established in this paper. Description of the modifications and their justifications are provided below for each part of the algorithm.

5.2.1 Parallelization of the Inner-Loop Step and Optimal Increments

As shown in Prop. 5.1, the increment obtained after k iterations of the CG algorithm in a given inner-loop approximates the rank- k optimal posterior mean update $\mathbf{x}_{\Pi_{\text{dof}}}^a$ associated with the linear Bayesian problem defined by the quadratic cost function (92). This stems from the fact that the CG procedure minimizes the cost function in a Krylov subspace that approximately spans the leading k eigenvectors of the prior-preconditioned Hessian, $\hat{\mathbf{H}}_p$. The eigendecomposition of $\hat{\mathbf{H}}_p$ $\{(\mathbf{v}_i, \lambda_i), i = 1, \dots, k\}$ is the basis of the two optimal posterior mean updates $\mathbf{x}_{\Pi_{\text{dof}}}^a$ (40) and $\mathbf{x}_{\text{FR}_{\text{dof}}}^a$ (68). Therefore, for each inner-loop step, either the Ritz pairs obtained from a CG minimization or the approximation of $\{(\mathbf{v}_i, \lambda_i), i = 1, \dots, k\}$ computed from a randomized SVD method can be used in combination with formula (40) or (68) to define optimal truncated posterior solutions. Assuming a large number of processors are available, our numerical experiments in Section 4 clearly suggest that replacing an iterative CG minimization algorithm by a parallel implementation of a randomized SVD computation would dramatically enhance the computational efficiency of the 4D-Var algorithm. In particular, in an operational context where only ~ 10 inner-loop iterations can be afforded for each CG minimizations, a randomized SVD algorithm is expected to provide many more approximated eigenpairs of $\hat{\mathbf{H}}_p$ than the Ritz pairs. Note that in practice it may be necessary to oversample $\hat{\mathbf{H}}_p$ in order to obtain accurate SVD estimates (see 3.1). As discussed in Halko *et al.* (2011), for most problems an oversampling parameter of ~ 10 is sufficient to obtain satisfying results, so that the number of required parallel integrations of the tangent-linear and adjoint models is roughly equal to number of eigenpairs of $\hat{\mathbf{H}}_p$ one wants to approximate. For the sake of simplicity, we shall use the term inner-loop to describe any quadratic minimization step of the incremental 4D-Var (i.e., any linearization step of the Gauss-Newton algorithm), even though in the context of randomized SVD methods the minimization is not performed through an iterative algorithm.

As discussed in 2.3.2, for a given inner-loop and a given approximation of k eigenpairs of $\hat{\mathbf{H}}_p$ $\{(\tilde{\mathbf{v}}_i, \tilde{\lambda}_i), i = 1, \dots, k\}$, an optimal approach to define the truncated solution of the quadratic minimization problem can be designed using the optimality results established in Proposition 2.19. More precisely, here we present a strategy, described in Alg. 2, that can be employed to statistically minimize the error in the approximated increments.

Algorithm 2 Adaptive Posterior Increment

For a given inner-loop, let $\{(\tilde{\mathbf{v}}_i, \tilde{\lambda}_i), i = 1, \dots, k\}$ be an approximation of the first k eigenpairs of $\hat{\mathbf{H}}_p$:

- 1: if $\tilde{\lambda}_k \leq 1$, then the approximated solution to the quadratic minimization problem (92) is:

$$\mathbf{x}_{\text{FR}_{\text{dof}}}^a = \mathbf{L} \left[\mathbf{Id} - \left(\sum_{i=1}^k \tilde{\lambda}_i (1 + \tilde{\lambda}_i)^{-1} \tilde{\mathbf{v}}_i \tilde{\mathbf{v}}_i^T \right) \right] \mathbf{L}^T \mathbf{H}^T \mathbf{R}^{-1} \mathbf{d}, \quad (111)$$

- 2: if $\tilde{\lambda}_k > 1$, then the approximated solution to the quadratic minimization problem (92) is:

$$\mathbf{x}_{\Pi_{\text{dof}}}^a = \mathbf{L} \left[\sum_{i=1}^k \left(1 + \tilde{\lambda}_i \right)^{-1} \tilde{\mathbf{v}}_i \tilde{\mathbf{v}}_i^T \right] \mathbf{L}^T \mathbf{H}^T \mathbf{R}^{-1} \mathbf{d}. \quad (112)$$

Based on the optimality results of Prop. 2.19, it is clear that when $\lambda_k \leq 1$ the choice $\mathbf{x}_{\text{FR}_{\text{dof}}}^a$ for the increment will always yield a smaller average approximation error than $\mathbf{x}_{\Pi_{\text{dof}}}^a$. However, the choice $\mathbf{x}_{\Pi_{\text{dof}}}^a$ when $\lambda_k > 1$ does not ensure a smaller average approximation error than $\mathbf{x}_{\text{FR}_{\text{dof}}}^a$. The rationale here is that the terms λ_i^3 in (71) will greatly dominate the terms λ_i in (74) as soon as

λ_i is significantly greater than 1. This threshold should therefore be adjusted when appropriate, depending on the typical spectra of $\hat{\mathbf{H}}_p$ for a given inverse problem. Indeed, as shown in our numerical experiments in Section 4.2 (Fig. 3 and Fig. 6), the approximations $\mathbf{x}_{\text{FRdof}}^a$ and $\mathbf{P}_{\hat{\mathbf{H}}_{\text{p.dof}}}^a$ can be associated with significantly lower approximation errors than $\mathbf{x}_{\text{p.dof}}^a$ and $\mathbf{P}_{\hat{\mathbf{H}}_{\text{p.dof}}}^a$, respectively, for a number of ranks k such that $k < \max\{j, \lambda_j > 1\}$. In the remaining we shall use a threshold of 1 for the sake of simplicity, but in principle improved objective criteria to define this threshold based on both heuristic and theory can be derived. This is left for future work.

Remarks 5.2. Statistically, Algorithm 2 should improve the convergence rate of incremental 4D-Var using standard CG minimizations, since by Proposition 5.1 the latter always produces the increment $\mathbf{x}_{\text{p.dof}}^a$, which yields a greater approximation error than $\mathbf{x}_{\text{FRdof}}^a$ whenever the smallest approximated eigenvalue of $\hat{\mathbf{H}}_p$ verifies $\hat{\lambda}_k \leq 1$. Note that this method can be readily integrated into existing incremental 4D-Var systems based on CG minimizations by extracting the Ritz pairs $\{(\tilde{\mathbf{v}}_i, \tilde{\lambda}_i), i = 1, \dots, k\}$ and applying Algorithm 2 instead of using the CG solution.

Remarks 5.3. In incremental 4D-Var, a stopping criterion needs to be defined for each inner-loop of the algorithm. In many operational DA systems, the stopping criterion corresponds to a fixed number of iterations defined by the available computing resources. Whenever possible, the choice can be improved, by, e.g., considering the absolute norm of the gradient of the cost function (92) (i.e., one wants $\|\nabla J^k\|_2 < \epsilon$, k being the iteration number), the relative change of the cost function (i.e., $|J^k - J^{k-1}| < \epsilon(1 + J^k)$), or a more sophisticated criterion based on sufficient conditions of convergence for Gauss-Newton minimizations, such as the one described in Lawless and Nichols [2006] (i.e., $\frac{\|\nabla J^k\|_2}{\|\nabla J^0\|_2} < \epsilon$). In the context where an approximation of k eigenpairs of $\hat{\mathbf{H}}_p$ $\{(\tilde{\mathbf{v}}_i, \tilde{\lambda}_i), i = 1, \dots, k\}$ is available and Algorithm 2 is used instead of the CG solution, the cost function J^k and its gradient ∇J^k at the optimal rank- k posterior mean can be evaluated around the values of the posterior means founds from Eqs. (111) and (112). When using the stopping criteria above in a randomized SVD context for instance, a minimum number of k iterations translates into a minimum number of $k + l$ samples (l being the oversampling parameter) for the random test matrix (see Algorithm 1). Alternatively, a stopping criterion that combines the deterministic convergence criteria of Lawless and Nichols [2006] and the statistical approach of Algorithm 2 can be defined as $k = \min\{i, \hat{\lambda}_i < 1 \text{ and } \frac{\|\nabla J^i\|_2}{\|\nabla J^0\|_2} < \epsilon\}$, since this guarantees both the convergence of the Gauss-Newton algorithm and the statistical optimality of the update $\mathbf{x}_{\text{FRdof}}^a$.

5.2.2 Optimal Posterior Perturbations in Ensemble-Based DA systems

In ensemble-based variational DA systems (e.g., EnVar), the propagation of errors from one assimilation window to the next is carried out using an ensemble of perturbed forecast trajectories that sample the posterior probability distribution. In practice, this posterior distribution needs to be evaluated at the maximum-likelihood (Tarantola, 2005). Therefore, a square-root of the posterior error covariance matrix evaluated at the last (converged) outer-loop iteration can be used to sample the posterior distribution. Two sampling strategies, one deterministic, and one stochastic, can then be adopted:

Stochastic Method The stochastic approach to sampling, similar to an EnKF, consists of generating an ensemble of random perturbations that sample the posterior distribution, using the formula:

$$\delta \mathbf{w}_i = \mathbf{S} \xi_i, i = 1, \dots, k, \quad (113)$$

where the ξ_i are independent random vectors drawn from a standard normal distribution $\mathcal{N}(0, 1)$, and \mathbf{S} is a square-root of the approximated posterior error covariance matrix.

Deterministic Method In the deterministic approach, similar to an EAKF, the ensemble of perturbations $\{\delta \mathbf{w}_i, i = 1, \dots, k\}$ is constructed so that it exactly verifies:

$$\mathbf{P}_{\text{approx}}^a = \delta \mathbf{W} \delta \mathbf{W}^T, \quad (114)$$

where $\mathbf{W} = \frac{1}{\sqrt{k-1}}(\delta \mathbf{w}_1, \dots, \delta \mathbf{w}_k)$ and $\mathbf{P}_{\text{approx}}^a$ is a given approximation of \mathbf{P}^a .

Based on Proposition 2.18, optimal rank- k posterior error covariance approximations can be constructed from the eigenpairs $\{(\mathbf{v}_i, \lambda_i), i = 1, \dots, k\}$ of $\widehat{\mathbf{H}}_p$. Given an estimate of the first k eigenpairs $\{(\tilde{\mathbf{v}}_i, \tilde{\lambda}_i), i = 1, \dots, k\}$ obtained, e.g., from iterative CG minimizations or randomized SVD techniques, a adaptive strategy to approximate the square-root of the posterior error covariance matrix is provided in Algorithm 3.

Algorithm 3 Adaptive Posterior Sampling

Let $\{(\tilde{\mathbf{v}}_i, \tilde{\lambda}_i), i = 1, \dots, k\}$ be an approximation of the first k eigenpairs of $\widehat{\mathbf{H}}_p$, computed at the last outer-loop iteration of an incremental 4D-Var optimization:

- 1: if $\tilde{\lambda}_k \leq 1$, then the approximated square-root for the posterior error covariance matrix is defined as:

$$\mathbf{S}_{\Pi_{\text{dof}}} = \mathbf{L} \left(\sum_{i=1}^k \left[(1 + \tilde{\lambda}_i)^{-1/2} - 1 \right] \tilde{\mathbf{v}}_i \tilde{\mathbf{v}}_i^T + \mathbf{Id} \right), \quad (115)$$

- 2: if $\tilde{\lambda}_k > 1$, then the approximated square-root for the posterior error covariance matrix is defined as:

$$\mathbf{S}_{\Pi_{\text{dof}}} = \mathbf{L} \sum_{i=1}^k (1 + \tilde{\lambda}_i)^{-1/2} \tilde{\mathbf{v}}_i. \quad (116)$$

Both Eq. (116) and (115) can be used with a stochastic method to generate optimal posterior perturbations. The approximated square-root (116) can also be used in a deterministic approach by setting $\delta \mathbf{w}_i = \sqrt{k-1} \mathbf{L}(1 + \lambda_i)^{-1/2} \mathbf{v}_i$ in (114). Note that the full-rank nature of (115) prevents it from being modeled as a deterministic square-root $\delta \mathbf{W}$ using a (necessarily) small number of perturbations $\delta \mathbf{w}_i$.

Remarks 5.4. The posterior square-roots (115) and (116) have also recently been proposed in Auligné *et al.* (2016) in the context of ensemble-variational DA for NWP. Their method (EVIL) uses the Ritz pairs obtained from the CG minimization at each inner-loop to construct the posterior square-root. The practical advantages of this approach and its better properties compared to other ensemble-based DA methods are discussed in Auligné *et al.* (2016). Our optimality results bring further theoretical justifications for using these posterior error square-root formulations, while providing new adaptive methods to construct the approximations for the posterior square-root and the posterior mean update. Additionally, Auligné *et al.* (2016) note that many Ritz pairs may be necessary to obtain an accurate posterior ensemble, which implies many iterations for the inner-loops and may not be practical for high-dimensional systems. It is further complicated by the fact that a rigorous approach would require one to use only the Ritz pairs obtained from the last inner-loop minimization, i.e., at the optimal state \mathbf{x}^a . This can have a significant impact when the system is highly non-linear, since in this case the Hessian matrix can be very different across inner-loops. Provided enough resources are available to perform many tangent-linear and adjoint integrations in parallel, the randomized SVD approach has the potential to approximate

many more eigenpairs of the prior-preconditioned Hessian than the CG technique. In practice, it would therefore provide a more accurate approximation of the square-root of the posterior error covariance matrix at the optimal state \mathbf{x}^a than obtained through the framework proposed by Auligné *et al.* (2016).

Remarks 5.5. Another possible way to define an approximated square-root of the posterior error covariance is to consider the preconditioner defined in Tshimanga *et al.* (2008), since their preconditioner effectively approximates the inverse Hessian of the 4D-Var cost function, which itself is an approximation of \mathbf{P}^a . In order to sample the posterior distribution, the preconditioner would be computed using only the approximated eigenpairs $\{(\tilde{\mathbf{v}}_i, \tilde{\lambda}_i), i = 1, \dots, k\}$ obtained at the last inner-loop, and the square-root constructed using the recursive formula provided in Thm. 2 of Tshimanga *et al.* (2008).

5.2.3 A New Randomized Incremental Optimal Technique (RIOT) for Variational Data Assimilation

Combining all the results from Section 5.2, a statistically optimal and computationally efficient ensemble-variational algorithm can be defined as follows:

Algorithm 4 One Cycle of the Randomized Incremental Optimal Technique for Variational Data Assimilation (RIOT VarDA)

Require: An initial time: t_0 ; a final time: t_f ; an initial prior vector: \mathbf{x}^b ; an initial innovation vector: $\mathbf{d} = \mathbf{y} - H(\mathbf{x}^b(t_0))$; a square-root of the prior error covariance matrix: \mathbf{L} ; a maximum number of outer-loop iterations (resource-dependent): m ; a maximum number of samples for the randomized SVD approximation (resource-dependent): r ; an oversampling parameter for the randomized SVD approximation: q ; a targeted rank for the randomized SVD approximation: $k = r - 2 - q$; a maximum number of samples for the non-linear trajectories (resource-dependent): p ³

- 1: **for** $i = 1, m$ **do**
 - 2: Compute $\mathbf{d} = \mathbf{y} - H(\mathbf{x}^b(t_0))$
 - 3: Using $(r - 2)$ samples for $\mathbf{\Omega}$ in Algorithm 1, generate \mathbf{Q} that approximates the range of $\widehat{\mathbf{H}}_p = \mathbf{L}^T \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \mathbf{L}$ at $\mathbf{x}^b(t_0)$, and evaluate the probabilistic error bound using two error samples ($l = 2$) in Eq. (90)
 - 4: Using \mathbf{Q} from step 3, and Algorithm 1, compute the SVD of $\widehat{\mathbf{H}}_p = \mathbf{L}^T \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \mathbf{L}$
 - 5: Update $\mathbf{x}^b(t_0)$ using Algorithm 2 to compute the increment $\delta \mathbf{x}^a(t_0)$:
 $\mathbf{x}^b(t_0) \leftarrow \mathbf{x}^b(t_0) + \delta \mathbf{x}^a(t_0)$
 - 6: **end for**
 - 7: Propagate $\mathbf{x}^b(t_0)$ to t_f using:
 $\mathbf{x}^b(t_f) \leftarrow H(\mathbf{x}^b(t_0))$
 - 8: Use the eigenpair approximations $\{(\tilde{\mathbf{v}}_j, \tilde{\lambda}_j), j = 1, \dots, k\}$ from outer-loop m to define an optimal posterior error square-root \mathbf{S} using Algorithm 3
 - 9: Construct p stochastic perturbations $\{\delta \mathbf{w}_i(t_0), i = 1, \dots, p\}$ at initial time t_0 using:
 $\delta \mathbf{w}_i(t_0) \leftarrow \mathbf{S} \xi_i, \xi_i \sim \mathcal{N}(0, 1), i = 1, \dots, p$
 - 10: Propagate the perturbations $\delta \mathbf{w}_i(t_0)$ to t_f using:
 $\delta \mathbf{w}_i(t_f) \leftarrow H[\mathbf{x}^b(t_0) + \delta \mathbf{w}_i(t_0)] - H[\mathbf{x}^b(t_0)]$
 - 11: Update the prior square-root \mathbf{L} using, e.g., the hybrid formulation of Eq. (94) with:
 $\mathbf{B}_{\text{ens}} = \mathbf{W} \mathbf{W}^T, \mathbf{W} = \frac{1}{\sqrt{p-1}}(\delta \mathbf{w}_1(t_f), \dots, \delta \mathbf{w}_p(t_f))$
-

³ All the resource-dependent parameters (i.e., m , r and p) should be defined from previous tests based on the particular application and computational resources available.

Remarks 5.6. An alternative strategy to using Alg. 1 is to use Algorithm 5.1 of Halko *et al.* (2011) to approximate the SVD of the square-root of the prior-preconditioned Hessian $\mathbf{U}_{\widehat{\mathbf{H}}_p} = \mathbf{L}^T \mathbf{H}^T \mathbf{R}^{-1/2} \approx \sum_{i=1}^k \sqrt{\lambda_i} \mathbf{v}_i \mathbf{w}_i^T$ (or $\mathbf{U}_{\widehat{\mathbf{H}}_p}^T = \mathbf{R}^{-1/2} \mathbf{H} \mathbf{L} \approx \sum_{i=1}^k \sqrt{\lambda_i} \mathbf{w}_i \mathbf{v}_i^T$). In Alg. 5.1, after application of the range finder (step 1 of Alg. 1), the operator $\mathbf{U}_{\widehat{\mathbf{H}}_p}$ (or $\mathbf{U}_{\widehat{\mathbf{H}}_p}^T$) needs to be projected onto the range \mathbf{Q} , i.e., one needs to perform the product $\mathbf{Q}^T \mathbf{U}_{\widehat{\mathbf{H}}_p} = (\mathbf{U}_{\widehat{\mathbf{H}}_p}^T \mathbf{Q})^T$, which requires k tangent linear integrations (or $\mathbf{Q}^T \mathbf{U}_{\widehat{\mathbf{H}}_p}^T = (\mathbf{U}_{\widehat{\mathbf{H}}_p} \mathbf{Q})^T$, which requires k adjoint integrations). Since the computational cost for both algorithms is largely dominated by the tangent linear and adjoint integrations, we see that Alg. 5.1 and Alg. 5.6 present similar complexities. Interestingly, if Alg. 5.1 is applied to $\mathbf{U}_{\widehat{\mathbf{H}}_p} = \mathbf{L}^T \mathbf{H}^T \mathbf{R}^{-1/2}$, one can use the random samples $\mathbf{H}^T \mathbf{R}^{-1/2} \boldsymbol{\Omega}$ generated as by-product of step 1 to estimate $\text{Im}(\mathbf{H}^T)$, since one has $\text{Im}(\mathbf{H}^T) = \text{Im}(\mathbf{H}^T \mathbf{R}^{-1/2})$. Therefore, Alg. 5.1 would enable a cost-free implementation of the representativeness error diagnostic (110).

Remarks 5.7. Although Algorithm 4 has been presented in the context of a strong-constrained 4D-Var system, it is applicable to EnVar systems by formally removing the explicit time dimension, which is now implicitly included in all vectors and operators. In particular, the increments $\overline{\delta \mathbf{x}}$, the 4D observational operator $\overline{\mathbf{H}}$ and the prior error covariance $\overline{\mathbf{B}}$ (represented by an ensemble) are now defined throughout the assimilation window (i.e., between t_0 and t_f). As a result, the tangent-linear and adjoint models include only spatial interpolation and measurement sensitivity operators (e.g., satellite averaging kernels), which are much easier to develop and faster to integrate than the tangent-linear and adjoint of the NWP model. In the context of EnVar the posterior error covariances are also defined by the updated (posterior) ensemble throughout the assimilation window, so that steps 7-11 of Alg. 4 are no longer needed. Instead, steps 2-5 can be applied to each prior perturbation member to generate the posterior ensemble. The sampling noise in the posterior ensemble can then be mitigated by using an objective 4D-Var localization method, such as the one proposed in Bocquet (2016).

6 Conclusion

This paper has presented a robust theoretical and numerical study of methods to approximate the solution of large-scale Bayesian problems. Our analysis focused principally on the construction of optimal low-rank projections for potentially high-dimensional Bayesian problems, taking into account the computational constraints of this framework, and in particular the need for matrix-free algorithms. A low-rank projection that maximizes the information content (i.e., the DOFS) of the inversion, the maximum-DOFS solution, was proposed and its optimality properties with respect to the posterior error covariance matrix and posterior mean approximations were analyzed in details. These results are also compared to other useful optimal low-rank approximations that are not associated with projections of the initial Bayesian problem. An interesting aspect of the low-rank approximations derived from the maximum-DOFS solution is that they allow one to adaptively optimize the posterior mean and the posterior error covariances based on the properties of the spectra of the so-called prior-preconditioned Hessian matrix.

The performance of the optimal projection and alternate low-rank approximations are tested in the context of an atmospheric source inversion problem whose dimension is small enough to allow exact computation of the posterior solution. Our results indicate good convergence properties for both the posterior error covariances and posterior mean approximations and are consistent with the theory. A large-scale version of this experiment is also presented, where the maximum-DOFS solutions are computed using an efficient randomized SVD algorithm, whose parallel implementation dramatically improves the scalability of the SVD computation upon standard iterative matrix-free methods (e.g., the BFGS or Lanczos algorithms).

Finally, we discussed the link between the maximum-DOFS low-rank approximation and the square-root formulation of incremental 4D-Var methods used in current operational DA systems. In particular, we showed the theoretical equivalence between k iterations of the preconditioned conjugate-gradient algorithm in inner-loop (quadratic) minimizations and the rank- k maximum-DOFS solutions of the associated projected Bayesian problem. Leveraging both our optimality results (e.g., the use of adaptive approximations in quadratic inner-loop minimizations) and the computational efficiency of the randomized SVD algorithms, we then proposed an improved implementation of incremental 4D-Var (RIOT). This approach is very generic and can be used with any square-root formulation of incremental 4D-Var, including hybrid ensemble-4D-Var (Clayton *et al.*, 2013).

Randomized SVD methods can be exploited to massively parallelize adjoint-based 4D-Var minimization algorithms, and as such represent an alternative approach that could rival the computational efficiency of ensemble-based DA, while preserving the full-rank nature of the variational formulation. However, adjoint-free ensemble-based approach such as EnKF or EnVar (Buehner *et al.*, 2010) still present the advantage that they do not require the development and maintenance of an adjoint model. Currently, randomization methods fundamentally rely on the availability of an adjoint model. One challenge ahead is to design matrix-free randomization methods that could perform SVD based on the forward model only. Additionally, preconditioning techniques such as the ones proposed in Tshimanga *et al.* (2008) could be applied to the randomized SVD method to improve the accuracy of the inner-loop minimizations. Therefore, future studies should focus on leveraging both parallelization and preconditioning methods to maximize the efficiency of randomized SVD techniques in incremental 4D-Var.

Acknowledgments This work was supported by the NASA GEO-CAPE Science Team grant NNX14AH02G and the NOAA grant NA14OAR4310136. This work utilized the Janus supercomputer, which is supported by the National Science Foundation (award number CNS-0821794), the University of Colorado Boulder, the University of Colorado Denver, and the National Center for Atmospheric Research. The Janus supercomputer is operated by the University of Colorado

Boulder.

Appendix: Useful Lemmas and Their Proofs

Lemma .1. *If \mathbf{A} , \mathbf{B} and $\mathbf{C} > 0$ are two $n \times n$ Hermitian non-negative definite matrices, and \mathbf{S} is a $n \times m$ matrix, one has the following properties:*

$$\mathbf{A} > \mathbf{B} \Leftrightarrow \mathbf{B}^{-1} > \mathbf{A}^{-1} \quad (117)$$

$$\mathbf{A} \geq \mathbf{B} \Rightarrow \mathbf{S}^T \mathbf{A} \mathbf{S} \geq \mathbf{S}^T \mathbf{B} \mathbf{S} \quad (118)$$

where \geq (respectively $>$) denotes the Löwner partial ordering (respectively strict) within the set of Hermitian non-negative definite matrices.

Proof. A proof of (117) is given in Horn and Johnson (2012) (Corollary 7.7.4). We recall it here for the reader's convenience. Let $\rho(\cdot)$ be the spectral radius of a matrix. One has $\mathbf{A} > \mathbf{B}$ if and only if $\rho(\mathbf{A}^{-1}\mathbf{B}) < 1$. Since for all \mathbf{M} , \mathbf{N} Hermitian matrices $\rho(\mathbf{MN}) = \rho(\mathbf{NM})$, one has $\rho(\mathbf{A}^{-1}\mathbf{B}) = \rho(\mathbf{BA}^{-1})$. (117) follows immediately.

By linearity, demonstrating (118) is equivalent to demonstrating $\mathbf{A} \geq 0 \Rightarrow \mathbf{S}^T \mathbf{A} \mathbf{S} \geq 0$. It is easily proven by considering $\mathbf{x} \neq 0$ and $\mathbf{y} = \mathbf{S}\mathbf{x}$. Since $\mathbf{A} \geq 0$, in particular $\mathbf{y}^T \mathbf{A} \mathbf{y} = \mathbf{x}^T \mathbf{S}^T \mathbf{A} \mathbf{S} \mathbf{x} \geq 0$, which shows that $\mathbf{S}^T \mathbf{A} \mathbf{S} \geq 0$. \square

The demonstration of Theorem (2.13) requires the following lemma (see Horn and Johnson (2012), Corollary 4.3.39):

Lemma .2. *Let \mathbf{A} be a $(n \times n)$ Hermitian matrix, suppose that $1 \leq p \leq n$, and let \mathbf{U} be a $(n \times p)$ matrix whose p columns form an orthonormal basis. Let $\lambda_1 \leq \dots \leq \lambda_n$ be the eigenvalues of \mathbf{A} . Then one has:*

$$\begin{aligned} \sum_{i=1}^p \lambda_i &\leq \text{Tr}(\mathbf{U}^T \mathbf{A} \mathbf{U}) \\ \sum_{i=n-p+1}^n \lambda_i &\geq \text{Tr}(\mathbf{U}^T \mathbf{A} \mathbf{U}) \end{aligned} \quad (119)$$

Moreover, for each $p = 0, \dots, n$, one has:

$$\begin{aligned} \text{Tr}(\mathbf{V}_{\min}^T \mathbf{A} \mathbf{V}_{\min}) &= \sum_{i=1}^p \lambda_i \\ \text{Tr}(\mathbf{V}_{\max}^T \mathbf{A} \mathbf{V}_{\max}) &= \sum_{i=n-p+1}^n \lambda_i, \end{aligned} \quad (120)$$

where \mathbf{V}_{\min} (resp., \mathbf{V}_{\max}) represents the matrix whose columns are the singular vectors associated with the p smallest (resp., highest) singular values of \mathbf{A} .

Proof. This is a corollary of the so-called *Poincaré separation theorem*, which states that:

$$\lambda_i(\mathbf{A}) \leq \lambda_i(\mathbf{U}^T \mathbf{A} \mathbf{U}) \leq \lambda_{i+n-p}(\mathbf{A}), \forall i = 0, \dots, p, \quad (121)$$

where $\lambda_i(\mathbf{M})$ denotes the i th singular value of the matrix \mathbf{M} , the singular values being arranged in increasing order. To prove (119), one just needs to remark that the sum of the eigenvalues and the sum of the main diagonal elements of an Hermitian matrix are equal. (120) is trivial. \square

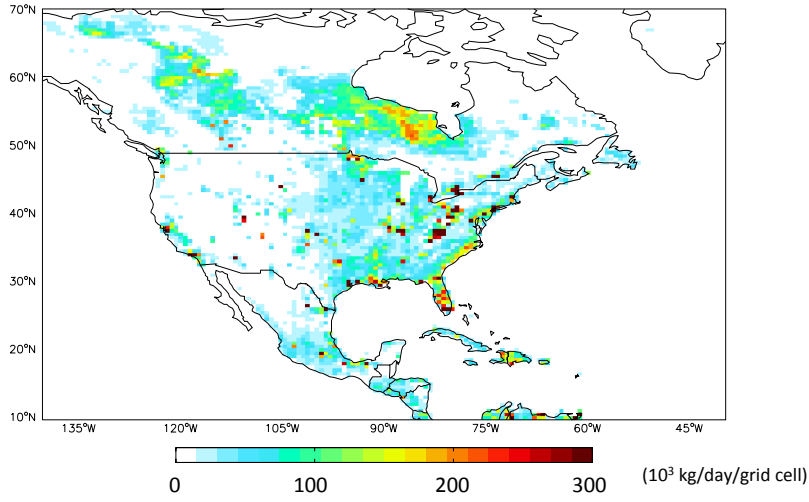


Figure 1: Monthly averaged total daily prior methane emissions for the nested North America domain ($0.5^\circ \times 0.7^\circ$). The size of the control vector for the original inverse problem is $n = 151 \times 121 = 18,271$. The reduced inverse problem considers only a subset of the grid-cells, resulting in a control vector of size $n = 300$.

Lemma .3. Let us define a matrix \mathbf{M} , a random vector \mathbf{q} with covariance matrix $\mathbb{E}(\mathbf{q}\mathbf{q}^T) = \mathbf{Q}$, and a Hermitian positive-definite matrix \mathbf{A} . Let us also define the square-roots $\mathbf{L}_\mathbf{Q}$ and $\mathbf{L}_\mathbf{A}$ of \mathbf{Q} and \mathbf{A} , respectively, i.e., $\mathbf{Q} = \mathbf{L}_\mathbf{Q}\mathbf{L}_\mathbf{Q}^T$ and $\mathbf{A} = \mathbf{L}_\mathbf{A}\mathbf{L}_\mathbf{A}^T$. One has:

$$\mathbb{E}\|\mathbf{Mq}\|_\mathbf{A}^2 = \|\mathbf{L}_\mathbf{A}^T \mathbf{M} \mathbf{L}_\mathbf{Q}\|_F^2 \quad (122)$$

Proof.

$$\begin{aligned} \mathbb{E}\|\mathbf{Mq}\|_\mathbf{A}^2 &= \mathbb{E}[(\mathbf{Mq})^T \mathbf{A} \mathbf{Mq}] \\ &= \mathbb{E}[\mathbf{q}^T \mathbf{M}^T \mathbf{L}_\mathbf{A} \mathbf{L}_\mathbf{A}^T \mathbf{Mq}] \\ &= \mathbb{E}[\text{Tr}(\mathbf{q}^T \mathbf{M}^T \mathbf{L}_\mathbf{A} \mathbf{L}_\mathbf{A}^T \mathbf{Mq})] \\ &= \mathbb{E}[\text{Tr}(\mathbf{L}_\mathbf{A}^T \mathbf{Mq} \mathbf{q}^T \mathbf{M}^T \mathbf{L}_\mathbf{A})] \\ &= \text{Tr}[\mathbf{L}_\mathbf{A}^T \mathbf{M} \mathbb{E}(\mathbf{q}\mathbf{q}^T) \mathbf{M}^T \mathbf{L}_\mathbf{A}] \\ &= \text{Tr}[\mathbf{L}_\mathbf{A}^T \mathbf{M} \mathbf{L}_\mathbf{Q} \mathbf{L}_\mathbf{Q}^T \mathbf{M}^T \mathbf{L}_\mathbf{A}] \\ &= \text{Tr}[\mathbf{L}_\mathbf{A}^T \mathbf{M} \mathbf{L}_\mathbf{Q} (\mathbf{L}_\mathbf{A}^T \mathbf{M} \mathbf{L}_\mathbf{Q})^T] \\ &= \|\mathbf{L}_\mathbf{A}^T \mathbf{M} \mathbf{L}_\mathbf{Q}\|_F^2 \end{aligned}$$

□

References

Anderson J. 2001. An ensemble adjustment Kalman filter for data assimilation. *Monthly Weather Review* **129**(12): 2884–2903.

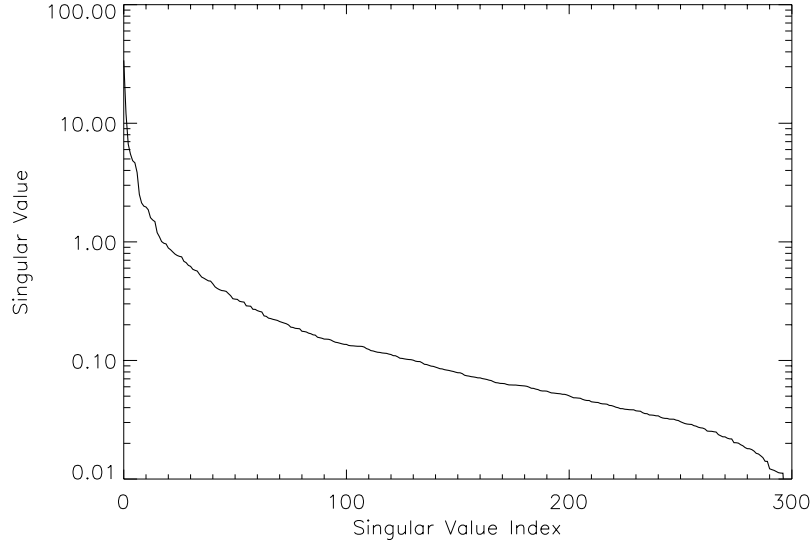


Figure 2: Singular value spectra of the prior-preconditioned Hessian matrix of the reduced source inversion problem.

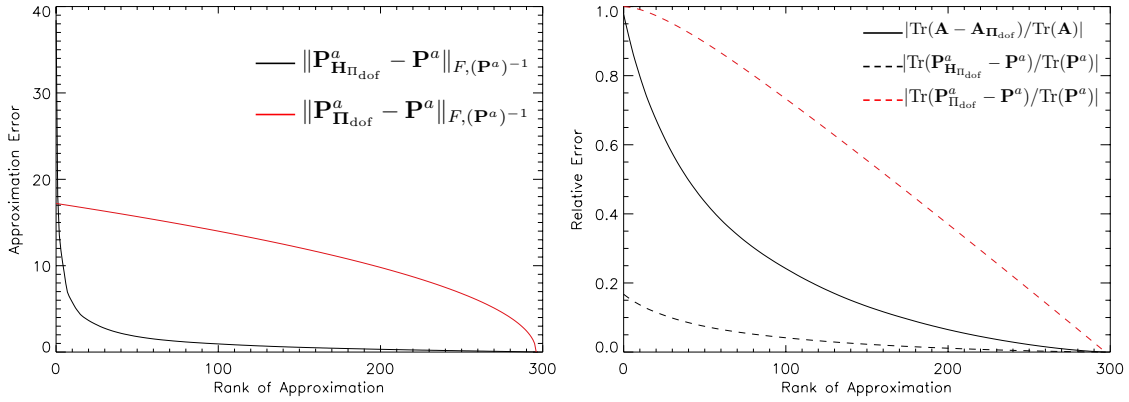


Figure 3: Left: \mathbf{P}^a -weighted error in the posterior error covariance matrix approximations $\mathbf{P}_{\mathbf{H}_{\Pi_{dof}}}^a$ (black line) and $\mathbf{P}_{\Pi_{dof}}^a$ (red line), as a function of the rank k of the approximation; Right: relative error in the DOFS approximation for solution of the maximum-DOFS projection ($\mathbf{A}_{\Pi_{dof}}$) (black solid line), and relative error in the total variance approximations $\text{Tr}(\mathbf{P}_{\mathbf{H}_{\Pi_{dof}}}^a)$ (black dashed line) and $\text{Tr}(\mathbf{P}_{\Pi_{dof}}^a)$ (red dashed line), as a function of the rank k of the approximation.

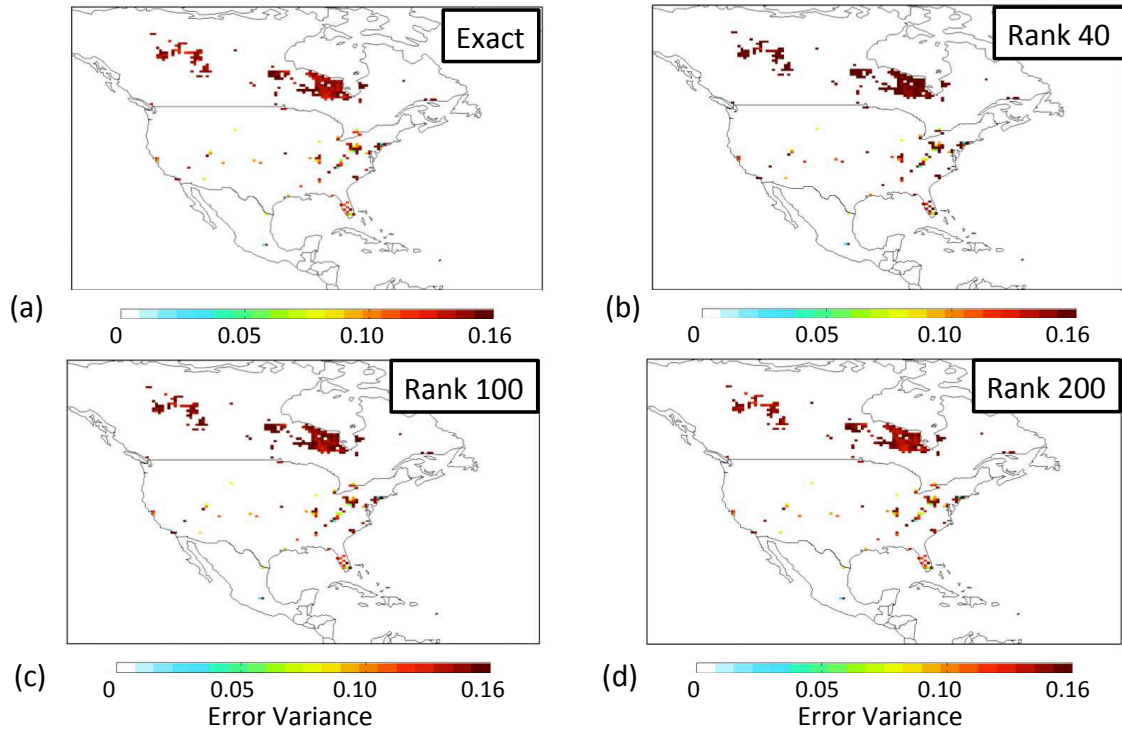


Figure 4: Exact and approximated posterior error variances using the low-rank update estimation $\mathbf{P}_{\mathbf{H}_{\Pi_{\text{dof}}}}^a$: (a) exact variances; (b) variances for a rank-40 update; (c) variances for a rank-100 update; (d) var

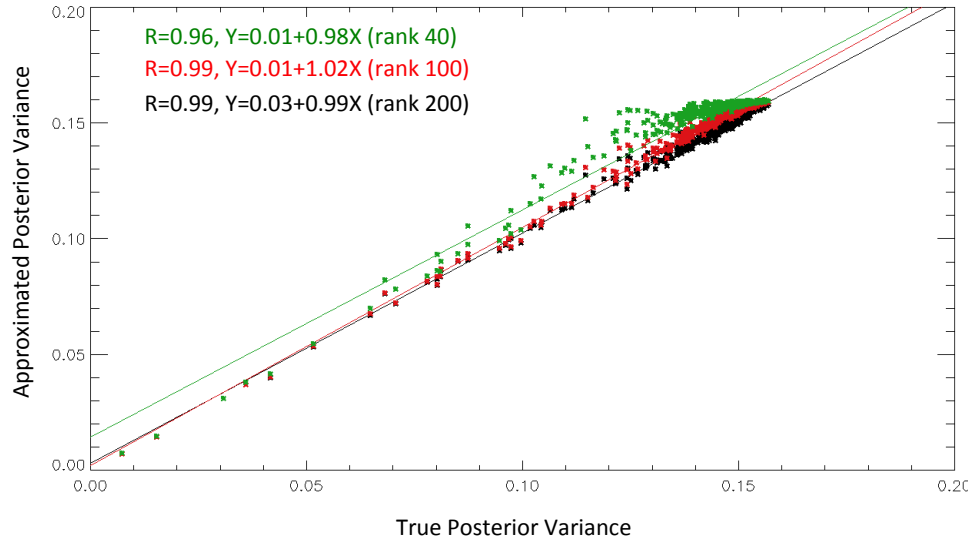


Figure 5: Scatterplot of true and approximated posterior error variances using the low-rank update estimation $\mathbf{P}_{\mathbf{H}_{\Pi_{\text{dof}}}}^a$, for different value of the rank k . Least-squares fit lines are shown along with the corresponding Pearson correlation coefficients (R) and linear fit equations ($Y=b+aX$). Green: variances for a rank-40 update; red: variances for a rank-100 update; black: variances for a rank-200 update.

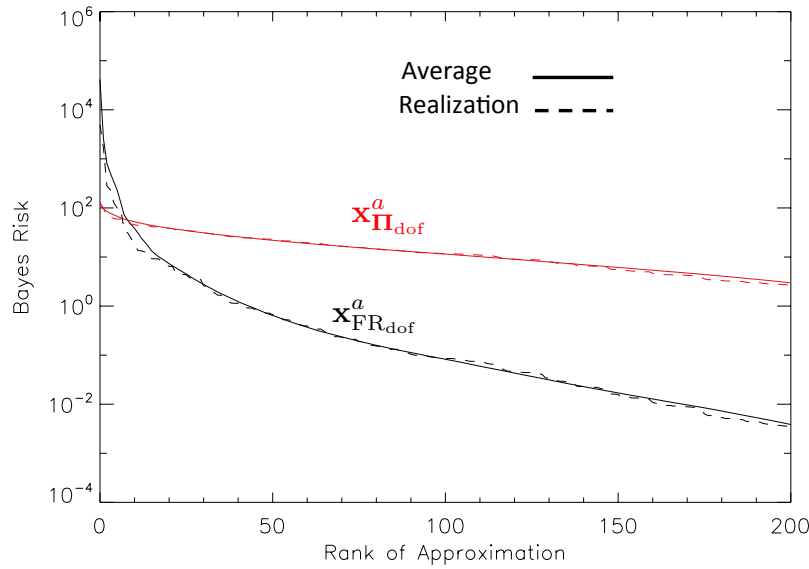


Figure 6: Average \mathbf{P}^a -weighted error in the posterior mean approximation (or Bayes risk), for the solution of the maximum-DOFS projection $\mathbf{x}_{\Pi_{dof}}^a$ (red solid line) and for the full-rank posterior mean approximation $\mathbf{x}_{FR_{dof}}^a$ (black solid line), as a function of the rank of the approximation. Results for one single realization of the prior and the observations are also shown in dashed lines for both approximations.

- Anderson J, Lei L. 2013. Empirical localization of observation impact in ensemble Kalman filters. *Monthly Weather Review* **141**(11): 4140–4153.
- Auligné T, Ménétrier B, Lorenc AC, Buehner M. 2016. Ensemble-variational integrated localized data assimilation. *Monthly Weather Review* **144**(10): 3677–3696.
- Bannister RN. 2008. A review of forecast error covariance statistics in atmospheric variational data assimilation. I : Characteristics and measurements of forecast error covariances. *Quarterly Journal of the Royal Meteorological Society* **134**(637): 1951–1970.
- Bennett AF. 2005. *Inverse modeling of the ocean and atmosphere*. Cambridge University Press.
- Bocquet M. 2016. Localization and the iterative ensemble Kalman smoother. *Quarterly Journal of the Royal Meteorological Society* **142**(695): 1075–1089.
- Bocquet M, Wu L, Chevallier F. 2011. Bayesian design of control space for optimal assimilation of observations. Part I: Consistent multiscale formalism. *Quarterly Journal of the Royal Meteorological Society* **137**(658): 1340–1356.
- Bousserez N, Henze D, Perkins A, Bowman K, Lee M, Liu J, Deng F, Jones D. 2015. Improved analysis-error covariance matrix for high-dimensional variational inversions: application to source estimation using a 3D atmospheric transport model. *Quarterly Journal of the Royal Meteorological Society* **141**(690): 1906–1921.
- Bousserez N, Henze DK, Rooney B, Perkins A, Wecht KJ, Turner AJ, Natraj V, Worden JR. 2016. Constraints on methane emissions in north america from future geostationary remote-sensing measurements. *Atmospheric Chemistry and Physics* **16**(10): 6175–6190.

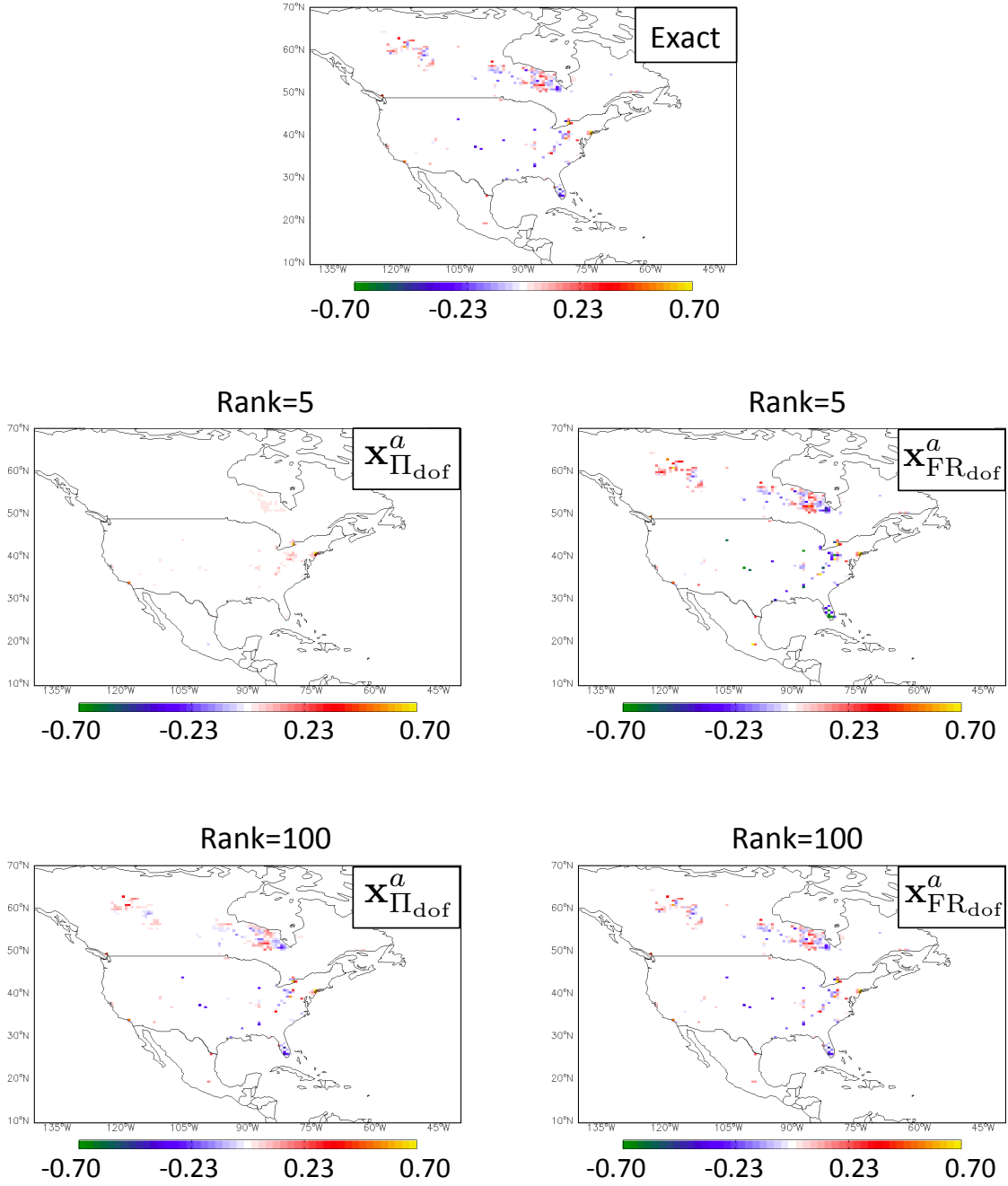


Figure 7: Exact (top) and approximated posterior flux increments for the solution of the maximum-DOFS projection $\mathbf{x}_{\Pi_{\text{dof}}}^a$ and the full-rank posterior increment approximation $\mathbf{x}_{\text{FR}_{\text{dof}}}^a$, for $k = 5$ and $k = 100$.

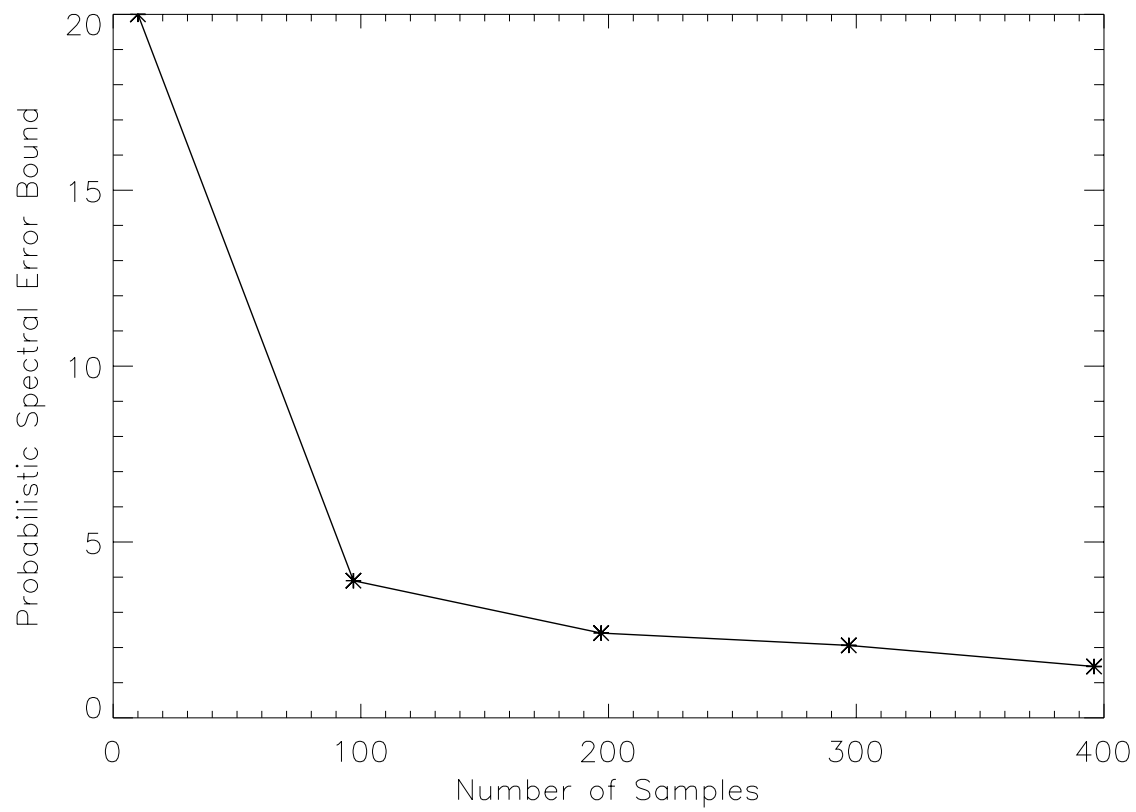


Figure 8: Probabilistic spectral error bound as a function of the number of samples used in the randomized SVD estimate.

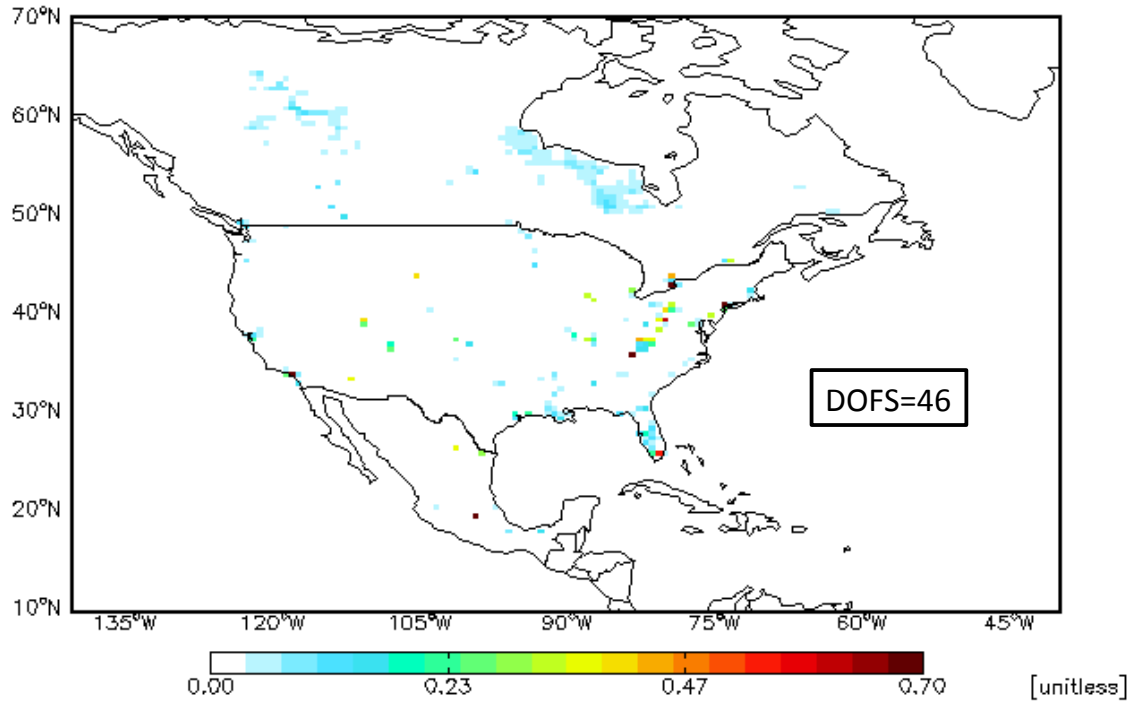


Figure 9: Diagonal of the model resolution matrix of the rank-400 maximum-DOFs projection ($\mathbf{A}_{\Pi_{\text{dof}}}$) for the full-dimensional inverse problem. Each element of the diagonal is associated with the observational constraints on the flux in one single grid-cell. The value quantifies the relative contribution (from 0 to 1) of the observations to the total information content, with respect to the prior information. The trace of the model resolution matrix, or DOFS, is also indicated.

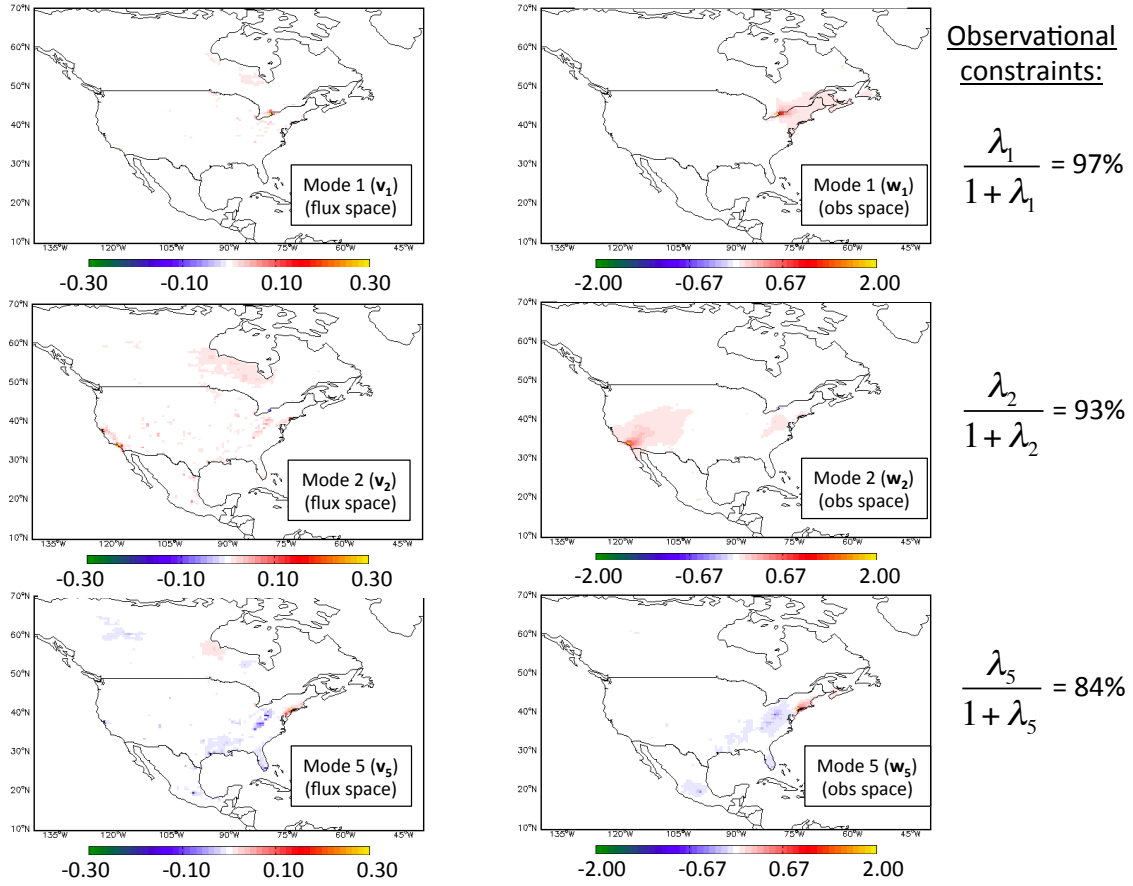


Figure 10: Three of the principal modes of the prior-preconditioned Hessian of the inversion ($\hat{\mathbf{H}}_p$), in control and observation space. Left panel: 1st, 2nd and 5th right singular vectors of the square-root of the prior-preconditioned Hessian $\hat{\mathbf{H}}_p^{1/2} \equiv \mathbf{R}^{-1/2} \mathbf{H}^T \mathbf{B}^{1/2}$. Right panel: 1st, 2nd and 5th left singular vectors of $\hat{\mathbf{H}}_p^{1/2}$. The relative contribution of the observations to the posterior information content (with respect to the prior) is also indicated on the right of the figures.

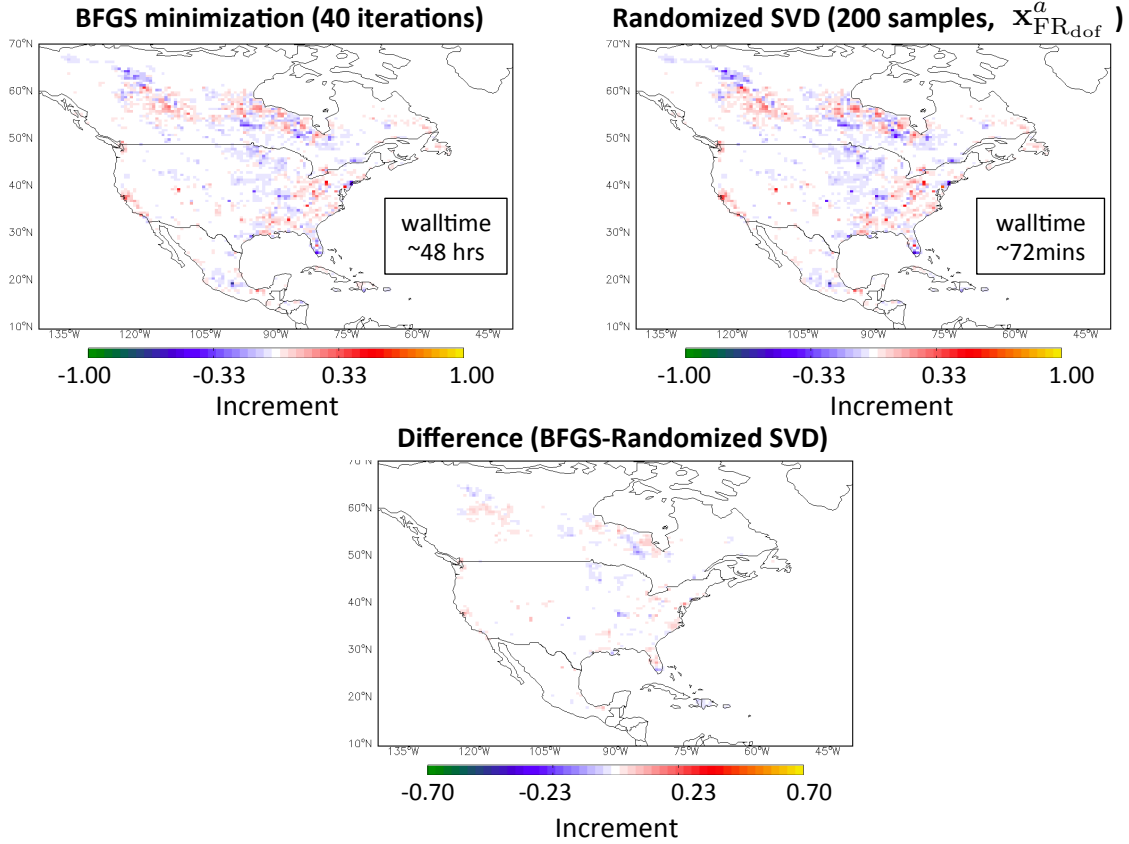


Figure 11: Comparison of the computational efficiency of a standard BFGS minimization with the adaptive approximation approach using a randomized SVD. Top left: posterior scaling factor flux increments after 40 iterations of the BFGS algorithm; Top right: posterior scaling factor flux increments for the adaptive approximation using a randomized SVD with 200 samples. In this case $\lambda_{200} < 1$, so the posterior increment is the full-rank approximation $\mathbf{x}_{\text{FR}_{\text{dof}}}^a$; Bottom: Difference between the posterior scaling factor increments obtained from the BFGS minimization and from the adaptive approximation using randomized SVD. The walltimes associated with the BFGS minimization and the adaptive approximation using randomized SVDs are also indicated on the top figures.

- Buehner M, Houtekamer P, Charette C, Mitchell HL, He B. 2010. Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP. Part I: Description and single-observation experiments. *Monthly Weather Review* **138**(5): 1550–1566.
- Bui-Thanh T, Burstedde C, Ghattas O, Martin J, Stadler G, Wilcox LC. 2012. Extreme-scale UQ for Bayesian inverse problems governed by PDEs. In: *Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis*. IEEE Computer Society Press, p. 3.
- Clayton A, Lorenc AC, Barker DM. 2013. Operational implementation of a hybrid ensemble/4D-var global data assimilation system at the Met Office. *Quarterly Journal of the Royal Meteorological Society* **139**(675): 1445–1461.
- Courtier P, Thepaut J, Hollingsworth A. 1994. A strategy for operational implementation of 4D-var, using an incremental approach. *Quarterly Journal of the Royal Meteorological Society* **120**(519): 1367–1387.
- Cui T, Martin J, Marzouk YM, Solonen A, Spantini A. 2014. Likelihood-informed dimension reduction for nonlinear inverse problems. *Inverse Problems* **30**(11): 114015.
- Desroziers G, Camino JT, Berre L. 2014. 4D-EnVar: link with 4D state formulation of variational assimilation and different possible implementations. *Quarterly Journal of the Royal Meteorological Society* **140**(684): 2097–2110.
- Friedland S, Torokhti A. 2007. Generalized rank-constrained matrix approximations. *SIAM Journal on Matrix Analysis and Applications* **29**(2): 656–659.
- Gaspari G, Cohn SE. 1999. Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society* **125**(554): 723–757.
- Golub GH, Van Loan CF. 2012. *Matrix computations*, vol. 3. JHU Press.
- Halko N, Martinsson PG, Tropp JA. 2011. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM review* **53**(2): 217–288.
- Henze DK, Hakami A, Seinfeld JH. 2007. Development of the adjoint of GEOS-Chem. *Atmospheric Chemistry and Physics* **7**(9): 2413–2433.
- Horn RA, Johnson CR. 2012. *Matrix analysis*. Cambridge university press.
- Isotalo J, Puntanen S, Styan GP. 2008. The BLUE’s covariance matrix revisited: A review. *Journal of Statistical Planning and Inference* **138**(9): 2722–2737.
- Jiang Z, Jones D, Kopacz M, Liu J, Henze DK, Heald C. 2011. Quantifying the impact of model errors on top-down estimates of carbon monoxide emissions using satellite observations. *Journal of Geophysical Research: Atmospheres* **116**(D15).
- Kopacz M, Jacob DJ, Henze DK, Heald CL, Streets DG, Zhang Q. 2009. Comparison of adjoint and analytical Bayesian inversion methods for constraining asian sources of carbon monoxide using satellite (MOPITT) measurements of CO columns. *Journal of Geophysical Research: Atmospheres* **114**(D4).
- Lanczos C. 1949. An iteration method for solving the eigenvalue problem of linear differential operators. *Bulletin of the American Mathematical Society* **55**(7): 717–718.
- Lorenc AC. 2003. The potential of the ensemble Kalman filter for NWP: A comparison with 4D-Var. *Quarterly Journal of the Royal Meteorological Society* **129**(595): 3183–3203.

- Ménétrier B, Auligné T. 2015. An overlooked issue of variational data assimilation. *Monthly Weather Review* **143**(10): 3925–3930.
- Ménétrier B, Montmerle T, Michel Y, Berre L. 2015. Linear filtering of sample covariances for ensemble-based data assimilation. Part I: Optimality criteria and application to variance filtering and covariance localization. *Monthly Weather Review* **143**(5): 1622–1643.
- Meurant G, Strakoš Z. 2006. The lanczos and conjugate gradient algorithms in finite precision arithmetic. *Acta Numerica* **15**: 471–542.
- Müller J, Stavrou T. 2005. Inversion of CO and NO_x emissions using the adjoint of the IMAGES model. *Atmospheric Chemistry and Physics* **5**: 1157–1186.
- Nocedal J, Wright S. 2006. Numerical optimization, second edition. *Numerical Optimization, Second Edition* : 1–664.
- Rabier F, Courtier P. 1992. 4-dimensional assimilation in the presence of baroclinic instability. *Quarterly Journal of the Royal Meteorological Society* **118**(506): 649–672.
- Rodgers CD. 2000. *Inverse methods for atmospheric sounding: theory and practice*, vol. 2. World scientific.
- Sherman J, Morrison WJ. 1949. Adjustment of an inverse matrix corresponding to a change in one element of a given matrix. *Annals of Mathematical Statistics* **20**: 317.
- Singh K, Jardak M, Sandu A, Bowman K, Lee M, Jones D. 2011. Construction of non-diagonal background error covariance matrices for global chemical data assimilation. *Geoscientific Model Development* **4**(2): 299–316.
- Spantini A, Solonen A, Cui T, Martin J, Tenorio L, Marzouk Y. 2015. Optimal low-rank approximations of Bayesian linear inverse problems. *SIAM Journal on Scientific Computing* **37**(6): 2451–2487.
- Tarantola A. 2005. *Inverse problem theory and methods for model parameter estimation*. SIAM.
- Thacker WC. 1989. The role of the Hessian matrix in fitting models to measurements. *Journal of Geophysical Research: Oceans* **94**(C5): 6177–6196.
- Tshimanga J, Gratton S, Weaver AT, Sartenaer A. 2008. Limited-memory preconditioners, with application to incremental four-dimensional variational data assimilation. *Quarterly Journal of the Royal Meteorological Society* **134**(632): 751–769.
- Turner A, Jacob D. 2015. Balancing aggregation and smoothing errors in inverse models. *Atmospheric Chemistry and Physics* **15**(12): 7039–7048.
- Wecht KJ, Jacob DJ, Sulprizio MP, Santoni G, Wofsy SC, Parker R, Bösch H, Worden J. 2014. Spatially resolving methane emissions in California: constraints from the CalNex aircraft campaign and from present (GOSAT, TES) and future (TROPOMI, geostationary) satellite observations. *Atmospheric Chemistry and Physics* **14**(15): 8173–8184.
- Wells K, Millet D, Bousserez N, Henze D, Chaliyakunnel S, Griffis T, Luan Y, Dlugokencky E, Prinn R, O'Doherty S, *et al.* 2015. Simulation of atmospheric N₂O with GEOS-Chem and its adjoint: evaluation of observational constraints. *Geoscientific Model Development* **8**(10): 3179–3198.
- Xu X, Wang J, Henze DK, Qu W, Kopacz M. 2013. Constraints on aerosol sources using GEOS-Chem adjoint and MODIS radiances, and evaluation with multisensor (OMI, MISR) data. *Journal of Geophysical Research: Atmospheres* **118**(12): 6396–6413.